

Supply Chain Forecasting

Syntetos, A.; Babai, Z.; Boylan, J.; Kolasssa, S.; Nikolopoulos, K.

European Journal of Operational Research

DOI:

[10.1016/j.ejor.2015.11.010](https://doi.org/10.1016/j.ejor.2015.11.010)

Published: 07/07/2016

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Syntetos, A., Babai, Z., Boylan, J., Kolasssa, S., & Nikolopoulos, K. (2016). Supply Chain Forecasting: Theory, Practice, their Gap and the Future. *European Journal of Operational Research*, 252(1), 1-26. <https://doi.org/10.1016/j.ejor.2015.11.010>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Supply Chain Forecasting: Theory, Practice, their Gap and the Future

(Accepted by the *European Journal of Operational Research*)

Aris Syntetos^{1#}, Zied Babai², John Boylan³, Stephan Kolassa⁴, Konstantinos Nikolopoulos⁵

¹Cardiff University, UK; ²Kedge Business School, FR; ³Lancaster University, UK;

⁴SAP Switzerland, CH; ⁵Bangor University, UK

#Corresponding author: +44 (0)29 2087 6572, SyntetosA@cardiff.ac.uk

HIGHLIGHTS

The literature on supply chain forecasting is critically reviewed;

The process of involving the forecasting community towards that task is described;

Gaps between theory and practice are identified;

Data and software related issues are explicitly considered;

Challenges are summarised followed by suggestions for further research.

ABSTRACT

Supply Chain Forecasting (SCF) goes beyond the operational task of extrapolating demand requirements at one echelon. It involves complex issues such as supply chain coordination and sharing of information between multiple stakeholders. Academic research in SCF has tended to neglect some issues that are important in practice. In areas of practical relevance, sound theoretical developments have rarely been translated into operational solutions or integrated in state-of-the-art decision support systems. Furthermore, many experience-driven heuristics are increasingly used in everyday business practices. These heuristics are not supported by substantive scientific evidence; however, they are sometimes very hard to outperform. This can be attributed to the robustness of these simple and practical solutions such as aggregation approaches for example (across time, customers and products).

This paper provides a comprehensive review of the literature and aims at bridging the gap between the theory and practice in the existing knowledge base in SCF. We highlight the most promising approaches and suggest their integration in forecasting support systems. We discuss the current challenges both from a research and practitioner perspective and provide a research and application agenda for further work in this area. Finally, we make a contribution in the methodology underlying the preparation of review articles by means of involving the forecasting community in the process of deciding both the content and structure of this paper.

Keywords: Supply chain forecasting; Forecasting software; Forecasting empirical research; Literature review

1. INTRODUCTION: THE NATURE OF SUPPLY CHAINS

A supply chain consists of all the parties involved, directly or indirectly, in fulfilling a customer request / demand (Chopra and Meindl, 2010). A 'party' is any decision making unit within the supply chain. It could be an organisation or a business unit within an organisation. The supply chain extends from the final customer through a variety of retailers, wholesalers and distributors, and goes back to the manufacturers and their component and raw material suppliers. Within the chain, there are flows of materials and products, information and money. Whilst financial flows are undoubtedly important, the focus in this paper is on the flows of materials, products and information. The integration of financial forecasts into an organisation's planning system is beyond the scope of this review.

The final customer's demand sets the entire supply chain in motion. It generates a course of actions at retailing organisations to respond to such demand, by having the necessary products and services in place to satisfy the customers. These ultimately involve the generation of requests / demand at the next level upstream ^(NOTE 1) in the supply chain, at wholesalers or distributors, who subsequently respond by placing requests on manufacturers, and so on. This upstream flow of requests constitutes the transmission of information from one supply chain member to another. This information flow is complemented by a flow of materials / products downstream the supply chain to satisfy these requests. Although the length of supply chains may vary considerably, satisfaction of the final consumers' requests is the *raison d'être* of all supply chains.

In addition to the *length* of supply chains discussed above (and the flow of materials and information across them) there is another key operational dimension involved in their structure: their *depth*. Supply chains are often geographically dispersed and they are sometimes referred to as supply *nets* rather than *chains* although we will retain the latter term for the purposes of our discussion. At any supply chain level, they involve many customers and suppliers (being placed in various locations) and of course very many products, all of which form, either separately or in combination with each other, natural *cross-sectional* hierarchies. Demand will then need to be aggregated at these various hierarchical echelons to inform decision making at a wide range of organisational and functional levels.

If the final consumers' demand were constant, or known with certainty well in advance, then the operation of a supply chain would be a straightforward (backwards) scheduling exercise. However, demand is not known and thus it needs to be forecasted. It is the uncertainty associated with this demand that makes supply chain management very difficult. In addition, the frequency with which forecasts are produced varies considerably not only between the various supply chain organisations but also within each of those organisations depending on the decision making process they serve. Retail inventory replenishments, for example, rely upon frequent short term forecasts, whereas aggregate sales planning may take place quarterly. This leads to another hierarchical feature of supply chains which is *temporal*, rather than cross-sectional, in nature.

The objective of every supply chain should be to maximise the overall value generated (Chopra and Meindl, 2010). The value (also known as supply chain surplus) a supply chain generates is the difference between what the final product is worth to the customer and the costs the supply chain incurs in filling the customers' requests. (These costs may be purely organisational or may include environmental costs as well). Such costs are an increasing function of the uncertainty associated with the demand and thus supply chain forecasting plays a major role in increasing the overall value.

Given the length of a supply chain which may be considerable, spanning various supplying levels, as well as the nature of supply chain decisions, that typically are hierarchical in nature, we argue that supply chain forecasting does not merely relate to specific techniques but rather to approaches and strategies that may capture these very supply chain characteristics. The purpose of this paper is *not* to provide the state of the art of forecasting methods ^(NOTE 2), unless the development of such methods originates in a supply chain context. Rather we emphasise forecasting strategies and approaches that stem directly from the structure and nature of supply chains. Is there a supply chain feature that gave birth to a particular forecasting method or makes a forecasting approach different in that context of application? If the answer is *yes*, then the relevant methods and approaches are discussed. If the answer is *no*, then references are given to other recent review and state of the art papers for interested readers to follow. The same principle applies also to reviewing the measurement of forecast performance. Forecast accuracy metrics and related advances in research are not covered in this paper as such metrics have not been developed from a supply chain perspective. However, we do refer to accuracy implication metrics (i.e. assessment of the utility rather than accuracy of the forecasts) as such metrics are of direct relevance to a supply chain setting.

At this point we should also mention that although the term ‘demand’ is often used in this paper when referring to forecasting, typically demand will not be known and actual sales are being used as an approximation. The terms ‘demand’ and ‘sales’ are used interchangeably in this paper, although strictly speaking the latter is often used as an approximation for the former.

In the next sub-section we discuss the main features of supply chains relevant to forecasting and we construct a framework to facilitate the conceptual positioning of various studies in the area of supply chain forecasting. We then discuss the methodological approach taken in this paper towards the organisation of the material and the structure and content of this work and we close with an indication of the organisation of the remainder of the paper.

1.1. The physiology of supply chains: length

The longer the supply chains are, and the more organisations they involve, the more difficult it becomes to coordinate them. The collaborative practices within supply chains vary considerably. There are three key features that have implications for supply chain forecasting:

- Under certain conditions, the variance of demand is amplified as it progresses upstream, making it more difficult to forecast accurately;
- There are potential gains in forecast accuracy which may be achieved by different forms of collaboration, including sharing of demand information between different levels of the supply chain;
- The practice of collaboration has resulted in some major initiatives like Collaborative Planning, Forecasting and Replenishment (CPFR) and Vendor Managed Inventory (VMI) systems that have had important implications for the practice of supply chain forecasting.

In Figure 2 we present graphically the first two issues discussed above. Please note that within each echelon (i.e. Supplier, Manufacturer and Retailer) there may be many such organisations. For example, a manufacturer’s supply chain may involve many suppliers and retailers, and a retailer’s supply chain may involve many manufacturers and suppliers.

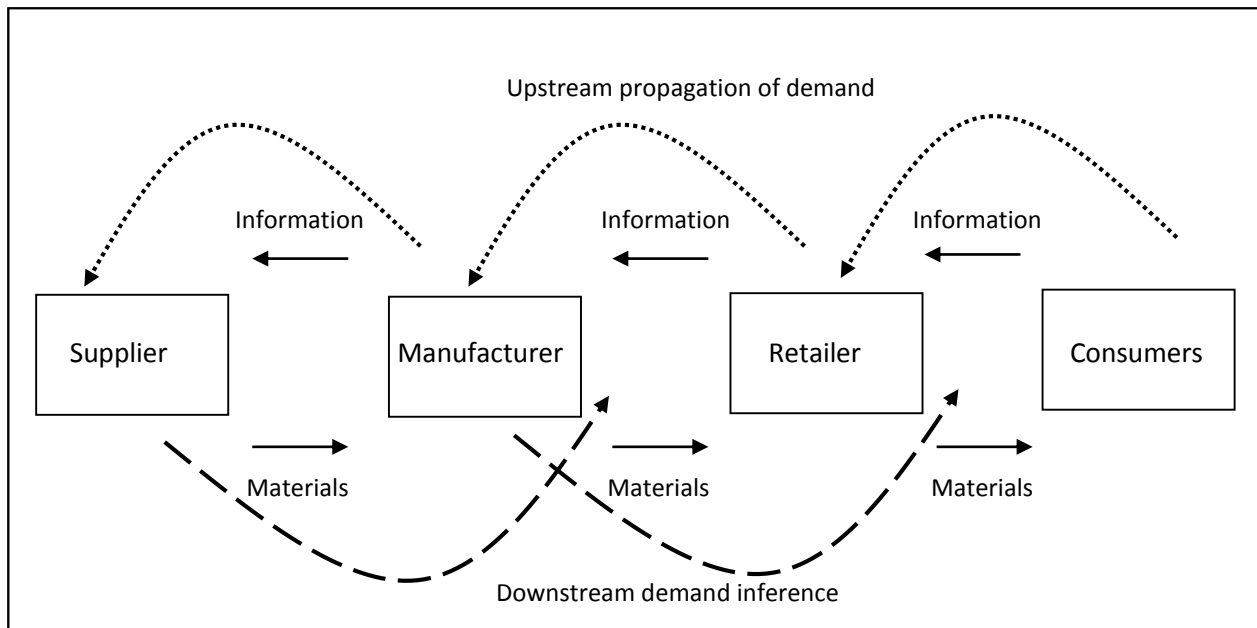


Figure 1. Supply chain structure: length

1.2. The physiology of supply chains: depth

Supply chain forecasting is a hierarchical process informing various levels of decision making. It spans from inventory control at the individual stock keeping unit (SKU) level to strategic planning at a highly aggregate level. A common implicit assumption in most of the operational research and operations management literature is that information is available at the required level of decision making; for example, lead time demand across all customers for a particular SKU in inventory control. However, such information is typically accumulated from order line transaction data or, indeed, it may be the outcome of distilling relevant information from higher order data. In general, the compatibility between the forecast output that is required at a certain decision making level and the input data, that form the time series based on which extrapolation is taking place, has received limited attention in the forecasting literature. Such hierarchical connections dominate real world practices and have significant implications for many aspects of supply chain decision making, from operational to strategic. Reconciliation of forecasts is an area of great importance and an integral part of Sales and Operations Planning (S&OP) processes.

There are some key features of supply chains that naturally lead to hierarchical structures:

- *Products:* Supply chains contain thousands of SKUs and decisions may be required, for example, at the individual SKU level in inventory control, at a product family level in Master Production Scheduling, or across all SKUs in aggregate capacity planning.
- *Suppliers:* Supply chains involve the supply of goods from a range of suppliers that may be as near as next door to their clients or as far away as the other side of the world. Similar geographical dispersion applies to distribution centres. Emphasis on specific suppliers may relate to the development of collaboration strategies and demand information sharing; or companies may be interested in a group of suppliers, that are geographically located close to each other, as part of a lead time reduction strategy from a specific part of the world; or they may be interested in all suppliers when it comes to a major decision on a new IT platform.

- *Customers*: Supply chains involve servicing a large number of industrial customers or final consumers that, further, will also be typically geographically dispersed. Companies will often be interested in the needs of a specific (big) customer in terms of prioritisation of activities, or in a group of customers that are geographically clustered together for transportation purposes, or in all customers for aggregate sales planning. Customers are frequently grouped into *markets*, which are groups of customers that share specific features. For instance, one may group a few key accounts into single-customer groups of their own, other large customers into one group, and medium to small customers in yet another group. Such a grouping into “markets” may reflect different go-to-market strategies or different marketing mixes, and forecasts may be desired at the aggregate market level.
- *Locations*: Suppliers and customers have in common that they may be geographically dispersed. In addition, a company that is forecasting for its supply chain will *itself* often have multiple locations to plan for, e.g., different factories and/or distribution centres.

1.3. The physiology of supply chains: time

In addition to the hierarchical effects discussed above, *time* may also lead to hierarchical structures in supply chains for the following reasons:

- *Time buckets*: The time buckets in which demand data are collected (that may also determine the forecast frequency) are rarely consistent across the companies forming a supply chain. This is not only a forecasting matter but also a reflection of the operational differences between the various stages of the supply chain. One stage might schedule activities by the day, another week by week and a plant might plan production monthly. The time buckets simply reflect the impact of these fundamental operational differences that make it very difficult to introduce the same bucket length for all stages;
- *Forecast horizons*: The forecast horizons involved in supply chain problems are very different. Inventory control, for example, necessitates forecasts over a lead time (or a lead time plus the review period, for periodic inventory control applications) (or in the presence of logistical constraints or complex ordering cost structures, which may offer opportunities for cost reductions by batching orders, forecasts may be needed for horizons longer than the lead time) whereas annual sales planning necessitates forecasts over a financial year;
- *History of the data*: Despite the recent advancements in IT, companies do not necessarily store long demand histories (and, depending on the circumstances, long histories may become obsolete quickly). Aggregation of demand data across (similar) SKUs may help to identify seasonal patterns, or other components that cannot be ‘seen’ at the individual SKU level due to the shortness of data. In addition, and as will be discussed later in the paper, the history of data available may also determine whether certain forms of temporal aggregation (non-overlapping) are feasible or not;
- *Frequency of demand*: Intermittence is a fundamental concept in supply chain forecasting. Sporadic demand characterises many supply chains and many forecasting methods have been proposed to handle such demands. These methods have not necessarily originated in the supply chain literature, but their properties do depend on supply chain features, as discussed

later in this paper. As previously discussed, in this work we do not focus on methods but rather on approaches and strategies, and such approaches will also be discussed for intermittent demand items;

- *Points in time*: For intermittent demand forecasting, not necessarily all points in time (all time buckets) are equally important for extrapolation purposes. Since replenishments are most often (but not always) triggered by a demand occurrence, the forecasts produced at the end of the demand occurring periods (i.e. when an issue occurs – *issue points*) are those that determine the inventory implications of using a particular estimator. The difference between *issue points only* and *all points in time* is an issue that is rooted in the supply chain context.

In the upper part of Figure 2 we present the hierarchical elements discussed in sub-sections 1.2 and 1.3 along three main dimensions: ‘location’, ‘products’ and ‘time’.

1.4 Theoretical framework

Thus far it has been argued that the hierarchical elements and the length of supply chains (the stages that they encompass) constitute their main features and determine the needs of supply chain forecasting.

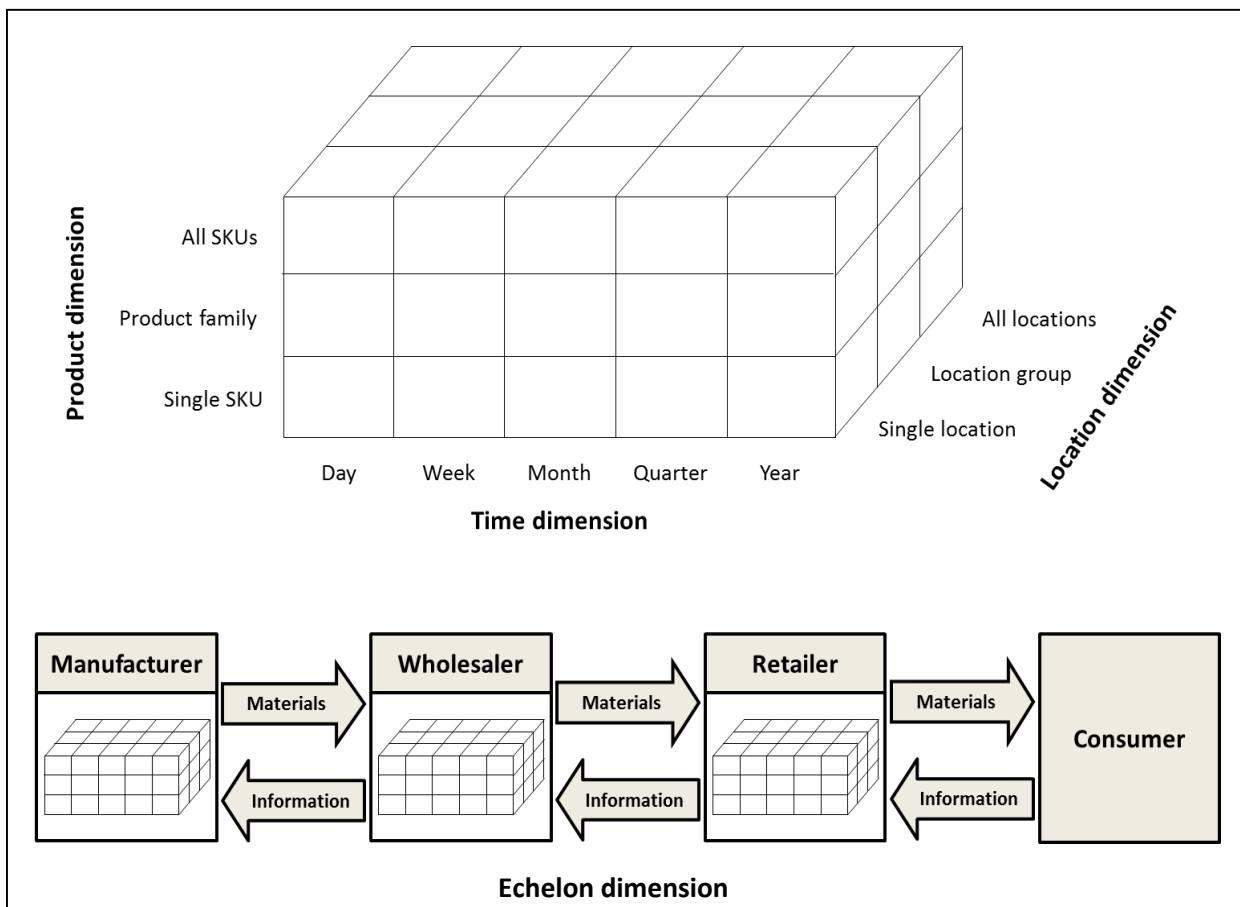


Figure 2. Supply chain structure: a framework

The framework presented in Figure 2 is informed by the requirements of supply chain forecasting and reflects its main dimensions (length – and the information and materials flow, and depth – and the pertinent hierarchical elements, including time). It will facilitate the organisation of this paper and the conceptual positioning of the studies we review.

The framework offers a four-dimensional structure within which supply chain forecasting hierarchies may be positioned. We claim that it is the simplest structure which can attain such positioning. The echelon dimension is necessary for any consideration of forecasting that relates to inventory management. The location dimension is also relevant to inventory management and is essential for any consideration of forecasts that inform transport planning. It is also crucial for inventory management/warehouse location decisions, as well as the decision to allocate given areas to different warehouses. The product dimension relates to inventory management, transport planning and also warehouse planning (e.g. where to locate products within a warehouse). Finally, the time dimension is essential for all forecasting problems, not just those that relate to supply chain forecasting.

1.5. Methodology: content and structure of the material

The structure of this paper and its content were determined based on a unique (to the best of our knowledge) approach that involved the contribution of the forecasting research community towards deciding on the thematic coverage of this work. We have tried to consult as many colleagues as possible and organise the paper around the areas that the forecasting community feels collectively are important and worth discussing in such a review paper rather than reflect our own background and particular perspectives.

To that end, a questionnaire, the aim of which was to identify the most important areas that should be addressed in this paper, was prepared and distributed through two main different routes:

1. First a hard copy of the questionnaire was distributed to the participants of the *26th European Conference on Operational Research (EURO)*, 1-4 July, 2013. The questionnaire may be found in *Appendix A* at the end of the paper.
2. Second, an announcement was made through the ORACLE (the news-magazine of the International Institute of Forecasters, IIF) inviting colleagues to electronically fill in the same questionnaire through *SurveyMonkey*.

All responses (in both cases) were anonymous, unless the participants wished to indicate their name in which case they participated in follow up discussions during the preparation of the manuscript. There were 43 completed questionnaires that were collected through the process(es) discussed above all of which were considered when structuring and writing the paper. The detailed analysis of the questionnaires can be found in *Appendix B*. (Any limitations in the paper remain of course attributable only to the authors of this work). The questionnaire results showed the following research areas as being the most important over the past 20 years: i) collaboration, ii) methods, and iii) planning. Those respondents who mentioned topics linked to ‘planning’ used terms such as ‘MRP’, ‘Lean’ and ‘Lifecycle Management’. Whilst these subjects are undoubtedly important, we considered them more relevant to ‘supply chain planning’ than ‘supply chain forecasting’ and so did not pursue them further. The theme of ‘collaboration’ was strongly linked to the term ‘information sharing’ and articles relating to collaboration and information sharing in forecasting are reviewed in Section 2 of this paper. The theme of ‘methods’ is very broad and includes such terms as ‘ARIMA’, ‘causal models’, ‘exponential

smoothing', and 'intermittent demand'. We have examined each of these topics in this review but found very little on causal methods in the context of supply chain forecasting and so this is picked up in the section on 'gaps in research'.

1.6. Organisation of the paper

The remainder of our paper is organised as follows: in the next three sections the main theoretical developments in the area of supply chain forecasting are reviewed. The organisation of the material follows the theoretical framework developed above to reflect the main dimensions of supply chain forecasting: *length* (discussed in Section 2), *depth* (discussed in Section 3) and *time* (discussed in Section 4). Each section concludes with a synthesis of the main gaps between the theory and practice of supply chain forecasting for the relevant dimension. Important analytical results in these areas are summarised and presented in three self-contained *Appendices* at the end of the paper (*Appendix C, D* and *E* for the issues of *length*, *depth* and *time* respectively). The summarisation of those results and the provision of a single point of reference to facilitate their collective retrieval is viewed as one of the contributions of this work. Section 5 addresses the role of judgment in supply chain forecasting. In Section 6 we review data and software related issues in the area of supply chain forecasting. Finally, in Section 7, we present a number of challenges for theorists, practitioners and software developers, and offer the conclusions of this work along with a number of suggestions for further research.

2. SUPPLY CHAIN FORECASTING: *LENGTH*

As a supply chain lengthens, there are more stages at which orders are placed. The original consumer demand is translated into an order from the retailer to replenish its stock; this depletes the stock at the next stage (e.g. wholesaler), which necessitates an order on the next stage (e.g. manufacturer), and so on until the end of the chain is reached.

The term "Bullwhip Effect" was coined by Lee et al. (1997) who observed amplification of demand variance through the stages of Procter and Gamble's supply chain for their product "Pampers" (babies' diapers). Lee et al. (1997) identified four causes of the Bullwhip Effect: demand signal processing, rationing and shortage gaming, batch ordering and price fluctuations. ^(NOTE 3)

In this section, we review research on the propagation of demand through the supply chain, and analytical results on the amplification of demand variance. We then proceed to the evaluation of Information Sharing as an approach to mitigating the Bullwhip Effect. Information sharing of the original consumer demand has been made possible by advances in Information Technology and by the adoption of such approaches as Collaborative Planning Forecasting and Replenishment (CPFR). There has been an ongoing debate in the academic literature on whether Information Sharing is necessary to counter the Bullwhip Effect. Some of the authors of this review paper have contributed towards this debate. Inevitably, this has coloured the line taken in this review, but we have endeavoured to give a fair summary of the views of those who disagree with us.

2.1. Upstream propagation of demand

Demand patterns may be described as stochastic processes, which represent the evolution of demand over time. For example, demand may be modelled as an Auto-Regressive Integrated Moving Average

(ARIMA) process. One of the simplest ARIMA processes is the AR(1) model, which captures first-order auto-regression of a demand series. This is an idealised demand process but there is some empirical evidence that such processes are adequate models for many demand series that are observed in practice. For example, Ali et al. (2012) found that 30% of the SKUs they investigated had demand that could be represented by AR(1) processes.

Lee et al. (2000) analysed the propagation of AR(1) demand for a two-stage supply chain, with one member at each stage of the chain. The downstream member places orders on the upstream member according to an Order-Up-To (OUT) inventory rule. The order-up-to-level is reviewed at the end of each period and amended in accordance with the revised forecasts. Lee et al. (2000) showed that an AR(1) demand process at the downstream members translates, using an OUT system, to an ARMA(1,1) order process on the upstream member. Their analysis of variances over lead-time identified three possibilities: i) if there is positive auto-correlation in the demand, then the variance of the orders is greater than the variance of the demand itself, and a Bullwhip Effect is observed; ii) if there is no auto-correlation, then there is no Bullwhip Effect; iii) if there is negative auto-correlation then there is an Anti-Bullwhip Effect, whereby the orders, over lead-time, are less variable than the demands over lead-time.

Lee et al.'s work on the propagation of demand through the supply chain, for an OUT inventory system, was later extended by Gilbert (2005) to a general $ARIMA(p, d, q)$ process. Gilbert (2005) showed that if downstream demand is of ARIMA form then, for an OUT system and an optimal (Minimum Mean Square Error) forecasting method, upstream demand will also be ARIMA. The upstream ARIMA process has the same orders of auto-regression and differencing as the downstream demand but has moving average terms of order $\max(p + d, q - L)$ for lead-time L . This establishes the dependence of the MA component of the ARIMA order process on the lead-time, as well as the demand process at the downstream member.

In practice, optimal forecasting methods are not always employed in supply chains. Often, simpler methods such as Simple Moving Averages (SMA) or Single Exponential Smoothing (SES) are used instead. Alwan et al. (2003) assumed that demand followed an $AR(1)$ process, but that forecasting is performed using SMA or SES. They proved that, for Simple Moving Averages of length n , an $AR(1)$ demand process with an OUT inventory system, translates to an $ARMA(1, n)$ process. For Single Exponential Smoothing, with smoothing parameter α , Alwan et al. (2003) proved that an $AR(1)$ demand process with an OUT inventory system translates to an $ARMA(1, \infty)$ process.

Ali and Boylan (2012) generalized the results of Alwan et al. (2003), for $AR(1)$, to an $ARIMA(p, d, q)$ demand process. They showed that, for an OUT system using Simple Moving Averages of length n , such a demand process translates to an $ARIMA(p, d, q + n)$ order process. The auto-regressive parameters remain unchanged; expressions for the upstream moving average parameters were also given by Ali and Boylan (2012).

Ali and Boylan (2012) generalized Alwan et al.'s results for Single Exponential Smoothing (SES). Instead of using an 'infinite representation' of SES (where the summation of weighted terms extend back to an infinite history), the authors adopted a more realistic finite representation, where the summation stops at t (the length of the data history). In this case, an $ARIMA(p, d, q)$ demand process translates, approximately, to an $ARIMA(p, d, t - 1)$ process.

In summary, considerable progress has been made over the last fifteen years in analysing the translation of demand processes through the supply chain, for an OUT inventory system. Complete representations have been found for $ARIMA(p, d, q)$ demand processes, both for optimal and non-optimal (SMA or SES) forecasting methods. These results are summarised in *Appendix C*. Ali (2008) showed how these results may be readily extended to multi-stage supply chains.

There are two restrictive assumptions that have been made in the research reviewed above: i) there is only one member at each stage of the chain; and ii) an OUT inventory system is employed. To make progress with the first issue, greater understanding is needed of cross-sectional aggregation of demand processes. This will be discussed in the next section of this review paper. To advance our understanding of the second issue, some simulation studies have been conducted on other inventory systems such as “order point, order-up-to-level”; see, for example, Syntetos et al. (2011). However, analytical results have not yet been obtained for such systems. A further limitation of the literature on demand propagation is the lack of empirical studies. There have been no studies, of which we are aware, comparing the upstream demand processes identified through standard criteria (e.g., AIC) with those expected from analytical results.

2.2. Demand information sharing

Lee et al. (2000) went further than simply analysing the propagation of demand processes through a supply chain. They also advocated “information sharing” of demand in supply chains, whereby the upstream member is informed not only the order quantity, Y_t , but also the demand on the downstream member, D_t . This enables the upstream member to forecast the downstream member’s demand directly, and eliminates reliance on the orders themselves. This would be beneficial, in terms of forecast accuracy, if the variance of demand is less than the variance of orders, as it is (for AR(1) demand) when there is positive auto-correlation in the demand series. This benefit in forecast accuracy flows through to inventory benefits at the upstream member of the supply chain.

There is some empirical research which supports Lee et al.’s conclusions, which were based on a theoretical analysis. Hosoda et al. (2008) analysed a real-world two-echelon (retailer-supplier) retail supply chain (focusing on a number of products that operate under an everyday low price strategy) and explored the benefit of the retailer sharing real Electronic Point Of Sales (EPOS) data with its supplier. The benefit of the supplier was assessed through the standard deviation of its forecast errors, which are linearly related to the inventory costs when safety stocks are determined based on the newsboy formulation. Standard deviation forecast error reductions between 8% and 19% were reported when exploiting the shared EPOS data. In addition, inventory benefits of demand information sharing have also been found for a multi-retailer, single manufacturer supply chain by Raghunathan (2003) (correlated demand between retailers) and by Cheng and Wu (2005) (uncorrelated demand).

Although these research results are encouraging, it must be acknowledged that empirical case-study evidence is scarce. This paucity of evidence is partly due to the slow take-up of Demand Information Sharing in practice and partly due to the lack of published evidence on it (although there is some case-study material, e.g., Barilla case, Hammond, 1994). Holweg et al. (2005) looked at several implementations of Collaborative Planning Forecasting & Replenishment (CPFR) and Vendor Managed Inventory (VMI) schemes across industries and countries, and concluded that the slow progress in adopting these strategies may be due to a lack of common understanding of such collaborative

concepts, in addition to the difficulty of integrating external collaboration with internal production and inventory control. Some of the characteristics required for successful initiatives were highlighted by Disney et al. (2001).

2.3. Downstream demand inference

Demand Information Sharing is valuable if it is possible to replace the task of forecasting a more variable process (the orders) with the task of forecasting a less variable process (the demand). This is the essence of the argument advanced by Lee et al. (2000) to support an Information Sharing strategy.

Lee et al. (2000) noted the following relationship, for AR(1) demand process of the form $D_t = \tau + \rho D_{t-1}$ and an OUT system, between the order process Y_t and the demand process D_t :

$$Y_t = D_t + \frac{\rho(1 - \rho^{L+1})}{1 - \rho} (D_t - D_{t-1}) \quad (1)$$

where ρ is the autoregressive parameter for the the AR(1) demand process and L is the lead-time.

They comment that, when the upstream member knows the parameters associated with the demand process D_t , the upstream member can utilize this equation to estimate the actual value of D_t . However, the authors proceed by assuming that the upstream member would not utilize this equation to infer the actual value of D_t .

Ragunathan (2001) develops this argument further. He re-writes the above equation by expressing D_t in terms of Y_t and D_{t-1} and uses this recursive relationship to obtain:

$$D_t = \frac{1 - \rho}{1 - \rho^{L+2}} \sum_{i=0}^{t-1} \left(\frac{\rho - \rho^{L+2}}{1 - \rho^{L+2}} \right)^i Y_{t-i} + \left(\frac{\rho - \rho^{L+2}}{1 - \rho^{L+2}} \right)^t D_0 \quad (2)$$

By substitution of equation (2) into equation (1), he obtains a recursive relationship for Y_t :

$$Y_t = \frac{1 - \rho^{L+2}}{1 - \rho} \tau + \frac{\rho^{L+2}(1 - \rho)}{1 - \rho^{L+2}} \sum_{i=0}^{t-1} \left(\frac{\rho - \rho^{L+2}}{1 - \rho^{L+2}} \right)^i Y_{t-i} + \rho^{L+2} \left(\frac{\rho - \rho^{L+2}}{1 - \rho^{L+2}} \right)^t D_0 + \frac{1 - \rho^{L+2}}{1 - \rho} \varepsilon_{t+1} \quad (3)$$

Therefore, Y_t can be estimated by the upstream member, based on the first two terms on the Right Hand Side of (3), as D_0 converges to zero, and ε_{t+1} , unknown to all supply chain members, is replaced by its expected value of zero. Of course, this depends on the upstream member having knowledge of the downstream member's demand stream parameters τ and ρ .

The variance of the expression, (3), converges to the variance from Demand Information Sharing as t approaches infinity. Raghunathan (2001) presents simulation results to show that the value of information sharing converges to zero rapidly. Hence, he concludes that there is no need to share demand information.

The above discussion is limited to AR(1) demand processes. However, analyses of ARMA(p,q) and ARIMA(p,d,q) processes by Zhang (2004) and Gilbert (2005), respectively, also concluded that inference of demand renders Demand Information Sharing redundant.

All of the papers reviewed in this section are based on the assumption of demand processes with unchanging parameters. There have been no analytical studies, to the best of our knowledge, which address models with time-varying parameters.

2.4. Downstream demand inference: counter-claims

Ali and Boylan (2012) revisited the issue of Downstream Demand Inference by questioning the assumption that the downstream member would share, for each individual Stock Keeping Unit, the specific ARIMA process (e.g., ARIMA(0,1,1)) and the specific ARIMA parameters but not share the demand value itself. This had been the common assumption made by Lee et al. (2000), Raghunathan (2001), Zhang (2004) and Gilbert (2005). Ali and Boylan argued that it is unlikely that supply chain partners would invest in a formal information sharing mechanism just to share the information on demand process and parameters and not the actual value of demand itself.

Ali and Boylan (2011) proved two 'Feasibility Principles', based on the assumption that an upstream member is not aware of the demand process or parameters at the downstream member. These principles apply for an ARIMA demand process with optimal (Minimum Mean Squared Error) forecasts.

Principle I: If the upstream member can infer the demand process at the downstream member, the demand values at the downstream member cannot be exactly calculated.

Principle II: If the upstream member cannot infer the demand process at the downstream member, then it may be possible for demand values at the downstream member to be calculated if a certain demand process is assumed from a restricted subset of the possible processes.

Taken together, the two principles imply that it is not possible for the upstream member to infer the demand process *and* the demand values at the downstream member.

Ali and Boylan (2012) also analysed the feasibility of inferring demand processes and demand values for ARIMA demand and non-optimal forecasting methods. For Single Exponential Smoothing, such inference is not possible. However, Ali and Boylan (2012) showed that this inference is possible for Simple Moving Averages, if the length of the Simple Moving Average is known. In this case, there is a one-one correspondence between demand and order processes (no change to p and d, and the upstream value of q is found by adding n – the length of the moving average – to the downstream value of q).

2.5. Current gaps

Research into the upstream propagation of demand processes, downstream demand inference and the benefits of demand (forecast) information sharing for supply chain forecasting currently flourishes and, as discussed in this section, many important developments have been made in this area. However, research has been predominantly theoretical in nature and subject to very specific assumptions and demand process formulations. Given the importance of this area for real world practices and the

(slowly) increasing implementation of collaborative approaches to supply chain forecasting and decision making, further research into the following areas would appear to be merited.

So far, academic investigations have looked only at OUT systems (see, e.g., Ali and Boylan, 2010). An exception is the study conducted by Syntetos et al. (2011), although that work relied upon simulation (system dynamics) rather than mathematical analysis. Only OUT systems have been used so far in terms of theory development and despite the fact that such systems are often employed for stock control purposes, there are certainly many other alternatives, the analysis of which should offer valuable insights for practitioners. A further restriction is that demand processes are generally assumed to be known and stationary.

There has been no analysis in this area for slow or intermittently moving items. Such an investigation would necessitate an Integer ARMA (INARMA) rather than ARIMA underlying framework. The relationship between these two frameworks is considered in some detail in Section 4 of the paper.

There is a need for some rigorous quantification of Information Sharing benefits and their linkages to forecast accuracy (and inventory reduction) benefits. Some real world case evidence has been presented but there is a need for more rigorous analysis which would isolate the effect of Information Sharing from other improvements in inventory management. Those who have advocated Downstream Demand Inference (DDI) as an alternative to Information Sharing have yet to offer any case studies of successful implementation of DDI.

3. SUPPLY CHAIN FORECASTING: *DEPTH*

3.1. Demand input and forecast output data

Demand for products (or service parts) is realized at an individual order line level, meaning that a particular business customer or end consumer requests a certain number of items for a particular SKU at a particular point in time. Demand is then naturally aggregated by the supplier along three key dimensions: locations, products and time, to inform decision making at various organizational and functional levels. This forms the basis of forecasting and planning, i.e. the extrapolation of some specific level of requirements into various time horizons (Bacchetti and Saccani, 2012; Boylan and Syntetos, 2007).

In practice, different decisions require forecasts at different levels of aggregation. For example: budgeting requires annual demand forecasts across all SKUs and customers; inventory control requires lead time demand forecasts per SKU and location; contractual arrangements and price negotiation rely upon annual forecasts across all relevant SKUs per customer; distribution management and scheduling require weekly or monthly demand forecasts across SKUs per customer/location – or aggregation across customers in a geographical area. A common implicit assumption in most of the forecasting and operational research (operations management) literature is that we utilize data that are as aggregate or disaggregate as our required forecasting output. However, the degree of aggregation of the forecasting output does not necessarily need to match with the existing data structure (which may be more aggregate or more disaggregate than the forecasts driving decision making). The degree of aggregation of the forecasting output (i.e. the forecast we use to make decisions) is actually a function of the decision making problem forecasting tries to support. This should be taken as a given for the

forecasting process and algorithms. On the contrary, inputs to the forecasting process are very often driven by existing data structures defined by the software solutions being employed and relevant organizational practices. Although the two may indeed match sometimes, this is not necessarily the case. A further challenge arises when we wish to use the same forecast for multiple purposes. In this situation, there is a need to design a forecasting process that addresses the multiple needs of the organisation, and which produces forecasts that are consistent across different levels of a hierarchy. These themes are discussed further in sub-sections 3.3 and 3.4.

Apart from the cases when demand inputs are at the required output aggregation level, the following three scenarios will typically occur:

- Forecasts are required at some higher level than the demand input data, for one of the three key dimensions: locations ^(NOTE 4), products and time. In this case, either forecasts are produced at the input level and then are aggregated to the required output level or demand is aggregated at the required output level and then extrapolation takes place at the level under concern.
- Forecasts are required at some lower level than the demand input data, for one of the three key dimensions: locations, products and time. In this case, either forecasts are produced at the input level and then disaggregated to the required output level or demand is disaggregated at the required output level and then extrapolation takes place at the level under concern.
- Forecasts are required at some lower (higher) level than the demand input data for some dimension(s) but at a higher (lower) level for some other dimension(s).

Although these hierarchical connections dominate real world practices and have significant implications for all levels of decision making, from tactical to operational to strategic, very little guidance may be found in the academic literature in this area. In addition, software providers are also currently in great need to further support the design of their demand management solutions with conclusive empirical findings and insights (Singh, 2013). Studying the effects of different forms of aggregation (location, product and time), and their combination, in relation to relevant decision making levels, in a systematic format that allows linkages to be identified and generic insights to emerge is yet to be conducted in the academic literature.

For the remainder of this section we emphasize single forms of aggregation across products and locations. The issue of temporal aggregation is discussed separately in Section 4. Although originating in fields other than supply-chain forecasting (mainly in economics and time-series studies), hierarchical forecasting very naturally reflects important supply-chain characteristics and offers ample scope for the introduction of innovative forecasting methodologies. First, we consider the problem of forecasting seasonal demand and present some important developments in this area that rely upon product cross-sectional aggregation practices. Then we review generic developments in product and customer (location) cross-sectional aggregation theory, followed by a discussion on forecast reconciliation which is an important requirement of S&OP. We close with remaining gaps in the area of forecasting by cross-sectional aggregation.

3.2. Cross-sectional aggregation

Demand uncertainty is among the most important challenges facing modern supply chains. In response, companies often invest in new forecasting software or consider implementation of new forecasting methods. However, there is another option: utilizing the hierarchical characteristics of data. Doing so does not necessarily require changing the suite of methods available for the forecasting task. Hierarchies offer an important forecasting opportunity, enabling logical techniques of aggregation and disaggregation. Cross-sectional aggregation in particular sums demands across items or customers (locations) and all items are reported for the same time periods. Useful information may then be extracted from the aggregate series that would otherwise be potentially lost at the individual SKU level due to the shortness of data.

Consider the problem of forecasting seasonal (trended or untrended) demand for example from *short* demand histories. Accurately estimating seasonality often requires lengthy time series. (Hyndman and Kostenko, 2007, discuss minimum sample size requirements for common seasonal forecasting models). However, if we cross-sectionally aggregate demand data to a group (say, product family) level, we gain the longer data histories typically associated with that family, making it easier to determine the seasonal pattern of demand. For instance, different brands of ice cream will have a similar seasonality with a summer peak, which may not be easily detected for low-volume flavours but can be estimated at a group level and applied on the product level. Similarly, one can consider grouping locations instead of (or in addition to) products: retail stores in a given region or even across multiple regions may have similar patterns, e.g., all along the seashore. However, care must be taken, as appropriate groupings may differ among locations and groups: seashore and ski resort stores will exhibit very dissimilar seasonal patterns of sunscreen lotion sales.

Traditionally, the seasonal component is estimated from an item's own data history, resulting in the determination of individual seasonal indices (ISI). However, when the demand history is short (and the data are very noisy), ISI will not provide accurate forecasts. In some cases, such as when there are only two (or less) years of data available, ISI may be worse than non-seasonal methods even when the data are seasonal (Chen and Boylan, 2008). A possible answer is to use the knowledge of demand at some group level to improve forecasts at the individual level. This approach is usually referred to as group seasonal indices (GSI). With GSI, instead of finding seasonal indices at the item level, demand is aggregated across a group of items that are thought to share a common seasonal pattern and seasonal indices are found at the group level. These indices are then applied to the individual product level in order to extrapolate future individual product demands. There are two approaches for obtaining group seasonal indices:

1. Sum demand across the product group and then estimate the seasonal indices from this aggregate series. Such an approach was proposed by Withycombe (1989) (WGSi).
2. Calculate individual seasonal indices (ISI) for each item in the group but then average these ISIs across the whole group and use this average for every single item in the group (Dalhart, 1974) (DGSi). A weighted average also could be used, with faster moving products having a greater weight.

From a practical perspective, it is then important to decide whether to:

1. Use the ISI approach for all series (which is what traditionally companies do);

2. Use the GSI approach for all series – and if so, whether to use DGSi or WGSi;
3. Choose, for each series, whether to use ISI or GSI – again deciding between DGSi and WGSi.

The third strategy was investigated by Chen and Boylan (2007) for untrended data, who found that the choice between ISI and the WGSi depends on the Coefficient of Variation (CV) of the deseasonalised series. Specifically, WGSi is more accurate than ISI if $CV_{INDIVIDUAL} > CV_{AGGREGATE}$. So the condition for WGSi being more accurate is that the coefficient of variation of the deseasonalised individual series ($CV_{INDIVIDUAL}$) must be greater than the coefficient of variation of the deseasonalised aggregate series ($CV_{AGGREGATE}$). [A complete set of rules indicates when it is best to apply the ISI, WGSi and DGSi methods (Chen and Boylan, 2007).] Chen and Boylan (2008) showed that the use of the rule can result in accuracy benefits on real data.

Dekker et al. (2004) and Ouwehand et al. (2005) proposed an aggregation approach for forecasting seasonal trended data. Level and trend are updated at individual product level. Seasonal indices though are estimated from the aggregated demand data of all products within a product family. These seasonal indices are then used, in conjunction with the item level and trend estimates, to forecast at the individual product level using the Holt-Winters method.

A key assumption in the application of all grouping methods is that a grouping mechanism is available. How the group is formed though can have a large influence on the forecast accuracy results (Ouwehand et al., 2005). Some practitioners would prefer to use their organization's standard product groups, which of course lend themselves to easy aggregation from a database and ERP standpoint. However, a group of items that may be homogenous in product characteristics, brand or any other feature chosen to define a product hierarchy may not necessarily have the same seasonal patterns (Zotteri et al., 2005). Therefore, it may be advantageous to use different groupings for seasonal forecasting. Boylan et al. (2014) examined theoretically, and by means of experimentation with empirical data, how groups should be formed in order to use the GSI methods discussed above. They proposed a *K-means* clustering method which was found to constitute a viable alternative to a company's own grouping method or that may be used with confidence if a company lacks a grouping method. For a further discussion on recent developments on the application of product-group seasonal indices to seasonal products please refer to Mohammadipour et al. (2012). The issue of seasonal grouping is one which would appear to merit further theoretical and empirical investigations. Finally, we note that the issues identified here also apply to grouping approaches to estimating other quantities of interest where a single product's time series may not yield enough information such as promotion effects, price elasticity, effects of calendar events, lifecycle patterns, cross-elasticities with other products etc.

3.3. Cross-sectional aggregation theory

In this section we consider some important developments in hierarchical product aggregation (sub-section 3.3.1) and hierarchical customer/location based decisions (sub-section 3.3.2). Both sub-sections 3.3.1 and 3.3.2 only consider aggregation within a specific hierarchy. Although the different hierarchies do not differ *intrinsically* as to how they should be treated in aggregation, we do distinguish between product and customer/location related studies. From an inventory perspective, the aggregation of requirements across a group of products ordered from the same supplier is a very natural thing to do and such a process is supported by inventory software packages. Similarly, the

aggregation of forthcoming requirements from a specific customer, or the aggregation of demand across customers in a specific region is important for marketing and sales initiatives and is supported by customer relationship management (CRM) packages. This is a very significant qualification to be made since, and as will be discussed in the next section of the paper, temporal aggregation approaches are not typically supported by software solutions.

3.3.1. Hierarchical product aggregation

Two approaches have dominated the literature to deal with hierarchical aggregation forecasting, namely bottom-up and top-down.

- The bottom-up approach involves forecasting at the individual SKU level (lowest level of the hierarchy) and then, if needed, aggregating such forecasts to obtain an estimate at some higher level of the hierarchy (see Hyndman et al., 2011). If a forecast is required at the individual SKU level only, then aggregation becomes redundant.
- The top-down approach involves forecasting at an aggregate level and then, if needed, disaggregating the forecasts to lower hierarchical levels (see, e.g., Boylan, 2010). For example, one could multiply the aggregate demand forecast by the (forecast of the) ratio of the respective individual item demand to aggregate demand (called *proration* in the literature, see, e.g., Fliedner, 1999; Strijbosch et al., 2008) in order to estimate individual SKU requirements. Gross and Sohl (1990) demonstrated, in a large empirical study, that simple sample averages for the *proration* factor are very effective ^(NOTE 5). Obviously, if a forecast is required at the aggregate level only, then disaggregation becomes redundant.
- A third variant, middle-out, essentially combines these two approaches by forecasting on a middle level in the hierarchy and then aggregating bottom-up, as well as disaggregating top-down.

Discussions on the comparative performance of top-down and bottom-up approaches have been present in the literature for a long time (not necessarily in a supply chain context only) and, to a certain extent, they are still inconclusive. An important determinant of such comparative performance is the hierarchical level (aggregate or subaggregate) for which a forecast is needed. The middle-out approach is usually subsumed under the two other approaches – considered from “below”, middle-out is top-down, while considered from “above”, middle-out is bottom-up.

Some scholars advocate the top down approach (Grunfeld and Griliches, 1960; Barnea and Lakonishok, 1980; Gross and Sohl, 1990; Fliedner, 1999), whereas others disagree (Orcutt et al., 1968; Edwards and Orcutt, 1969; Dangerfield and Morris, 1992; Gordon et al., 1997; Weatherford et al., 2001) and, among other reasons, argue that the outperformance of bottom-up is due to the fact that information loss is substantial when aggregating time series; that is, the data is too aggregated to represent diverse demands ^(NOTE 6). Some additional arguments that are often put forward by practitioners against top-down approaches have been summarised by Dawson (2013):

- Individual products may be in different classes;
- The parameters used for disaggregation (usually historical percentages of items’ contribution to the group) are subject to sampling variability;
- It is not possible to estimate the standard deviation at the individual level from the standard deviation at the aggregate level.

Nevertheless, there are instances in supply chain forecasting when the top-down approach is clearly more appropriate than bottom-up methods. Boylan (2010) argues that if there is a change in policy, for example, with regards to pack sizes, then aggregate data (in suitable units of measurement) is a better guide to the future than previous sales of a particular pack size. In this case, the aggregate forecast may require disaggregation using judgmental estimates (rather than sample derived proportions as discussed above) on the effect of introducing new pack sizes. In addition, and as discussed in the previous sub-section, longer data histories make model identification and component estimation more reliable at the aggregate level. However, it is true that if forecasts are required at the individual SKU level and thus we need to disaggregate, strong assumptions are being made that the lower level series are following the same pattern(s) as the aggregate one, and those assumptions may be incorrect.

An alternative line of attack is to take a contingent approach and analyse the conditions under which one method produces more accurate forecasts than another (e.g., Shlifer and Wolff, 1979; Lutkepohl, 1984; Viswanathan et al., 200; Rostami-Tabar et al., 2015). Widiarta et al. (2008) evaluated top-down versus bottom-up forecasting in a production planning context for the purpose of estimating SKU level requirements. In their work, aggregate demand series were assumed to be composed of several correlated sub-aggregate components. Each such component was assumed to follow a stationary (first order univariate moving average) time series process, which is correlated over time. Finally, as is common in a production-planning environment, it was assumed that simple exponential smoothing is used to extrapolate requirements under both methods. The researchers showed that the performance of the two methods is nearly identical, regardless of the coefficient of correlation between the item demands, the items' proportion in the family demand and the coefficient of the serial correlation term of the demand process.

In a later work, and under the same assumptions, Widiarta et al. (2009) examined analytically the relative effectiveness of top-down and bottom-up approaches for forecasting aggregate rather than SKU level demand. The authors concluded that: *"there is no difference in the relative performance when the time series for all of the sub-aggregate components follow a first-order univariate moving average process with identical coefficients of the serial correlation term (p. 87)"*. They also performed a simulation study to examine the case of differing coefficients of the serial correlation term among the sub-aggregate components. It was found that performance differences *"...are relatively insignificant when the correlation between the sub-aggregate components is small or moderate (p. 87)."*

Sbrana and Silvestrini (2013) extended the work of Widiarta et al. (2009) by employing an unrestricted multivariate framework allowing for interdependency between its variables. They showed theoretically (and by means of simulation) that the previously identified similar performance of bottom-up and top-down holds even when the moving average parameters of the individual components are not identical. In addition, they showed that the relative forecasting accuracy of the two methods depends on the parametric structure of the underlying framework.

More recently, Rostami-Tabar et al. (2015) analysed theoretically and by means of simulation on theoretically generated data the relative performance of top-down and bottom-up for forecasting both aggregate and SKU level demand. The latter was assumed to follow a non-stationary ARIMA (0,1,1) demand process and exponential smoothing (which is optimal for the process under concern) was assumed to be employed for forecasting purposes. An important finding was that the forecast accuracy

improvements achieved by bottom-up and top-down for non-stationary demands are higher than those associated with stationary cases. The theoretical findings were validated through empirical analysis with data from a European superstore; this is an important aspect of the study under concern since no other empirical evidence has been put forward in the literature in this area.

As previously noted in this sub-section, discussions remain largely inconclusive as to which method (top-down or bottom-up) performs better under which situation (forecasting at the aggregate or disaggregate level) and expanding the empirical knowledge base in this area would be of a great benefit for real world practices. Similarly, and since hierarchical forecasting is a complicated procedure, analytical results are hard to obtain and currently there is scope for more theoretical contributions in this area (Strijbosch et al., 2008; Strijbosch and Moors, 2010).

3.3.2. Hierarchical customer or location-based decisions

Hierarchical customer or location-based forecasting does not differ from product hierarchical forecasting in terms of the main issues determining the comparative performance of the bottom-up and top-down approach. Performance appears to be conditioned on the forecasting methods being used, the modelling assumptions being made (for analytical type work), the desired level of aggregation of the output of the forecasting process, and the criteria employed for constructing a 'family' of components (however these components are defined: customers, locations).

Flidner and Lawrence (1995) suggested that the method to identify group candidates is not very important. They observed that sophisticated clustering techniques used to identify family members did not lead to better forecast performance than when families were randomly determined. However, Zotteri et al. (2005) showed that what is driving forecast accuracy is the choice of the appropriate level of aggregation (e.g. store level, distribution centre level or supply chain level). Forecasters do not actually choose the level of aggregation of the output of the forecasting process. But this does *not* mean that they have no discretion as to the choice of the proper level of aggregation for producing the forecasts. Forecasting with more aggregated data than required lowers sampling error but raises specification error (Zotteri et al., 2014). The former stems from using data that is inadequate for estimating the demand pattern, either in terms of quantity (limited size of the actual sample) and / or quality (one-off events that may distort the picture of what is happening in terms of demand). The latter refers to pooling data from different sources, leading essentially to combining apples and oranges. For example, different supermarkets have different intra-day seasonality of customer traffic, so aggregating data across stores may lead to an overestimation of demand in one store at a particular time while underestimating demand at another time.

On the contrary, using more granular data than the process requires lowers specification error, but increases sampling error. The best forecasting strategy then is the one that considers the trade-offs between sampling and specification error and minimises the sum of these errors. Zotteri and Kalchschmidt (2007) studied the problem of aggregation across locations by evaluating the above discussed components of the forecasting error under the assumption of stationary non-correlated demand (over time and across products and markets). They concluded that the appropriate level of aggregation depends on the size and homogeneity of the supply chain and suggested metrics that one may adopt to support the choice of the appropriate forecasting process, thus providing help to managers in defining the proper level of aggregation for a specific situation. Further, they suggested that it would be interesting to model clustering of stores as an intermediate alternative between a

very detailed and a very aggregate model like the one considered in their research (Caniato et al., 2005). This would help companies to identify the level of aggregation that better solves the trade-off between specification and sampling error. This would probably entail the modelling of both intra-cluster and inter-cluster heterogeneity.

There have been only a few studies considering the effects of hierarchical location-based decision making despite the importance of such decisions for supply chain management. The trade-off between sampling and specification error should govern the choice of the level of aggregation to produce forecasts, and more empirical insights into how this trade-off may be best determined would be very valuable both from an academic and practitioner perspective.

3.4. Reconciliation of forecasts

The previous subsections addressed three “standard” ways of forecasting hierarchical time series: bottom-up, top-down and middle-out. One completely different and very promising approach to hierarchical forecasting has recently been proposed (Athanasopoulos et al., 2009; Hyndman et al., 2011; Hyndman and Athanasopoulos, 2014). In this *optimal combination* approach, one first forecasts all time series in the hierarchy separately and independently, which yields base forecasts that will usually not satisfy the summation constraints given by the hierarchy. These separate base forecasts are then combined using a linear transformation, which can encode information about covariances between the time series, if available. The final forecasts are then, by definition, hierarchically consistent.

This approach has a number of conceptual and practical advantages over bottom-up, top-down and middle-out approaches:

- The *entire* base forecast vector is included in the reconciliation, not only one particular level in the hierarchy. Therefore, one can naturally use different forecasting techniques and model different dynamics for different hierarchical nodes.
- This property of the optimal combination approach may explain why it has been found to improve forecasts across the entire hierarchy, in contrast to bottom-up aggregation, which often is found to be better at lower levels but worse at higher ones, and to top-down disaggregation, which often yields good forecasts at higher and worse ones at lower levels.
- This approach naturally allows combining multiple hierarchies, such as product, geographic or customer hierarchies. In principle, bottom-up can do so, too, by forecasting the most fine-grained time series, e.g., sales of a single SKU to a single customer in a single location – but such time series will frequently be intermittent and hard to forecast. Conversely, it is unclear how top-down and middle-out approaches could deal with multiple hierarchies without some sort of reconciliation at all, since *a priori*, the top level forecasts in the different hierarchies are likely to be different.
- The optimal combination forecast also easily addresses weighted summation. For instance, one could include sales prices in the summation matrix to forecast revenues, or purchasing prices, to forecast cash requirements.

However, some open issues remain.

- The modified bottom level forecasts can in principle end up negative, which may be the optimal mathematical solution to the generalized least squares (GLS) problem, but is likely to be nonsensical in a supply chain setting without returns. To avoid this, one can use non-negativity constraints on the problem.
- Realistic hierarchies can involve thousands of products and hundreds of locations. The corresponding GLS may be too large to solve without special numerical methods. Hyndman et al. (2011, section 5) discuss some potential remedies, and Hyndman et al. (2014a) have recently proposed a fast recursive approach for specific hierarchy topologies.
- Further, and in relation to the point raised above, forecasts from all levels need to be produced (as opposed to a particular level which is the case for either bottom-up or top-down) and this introduces an extra level of complexity when applied in practice.
- Finally, the fact that forecasts from all levels need to be taken into account when producing the final (reconciled) forecasts may also prove to be problematic when judgmental adjustments are performed at only one level, say the most aggregated or the most disaggregated one. Linking judgmental interventions to hierarchical forecasting is an issue that requires further research and this is discussed in more detail later on in the paper.

The optimal combination approach has been implemented in the *hts* package (Hyndman et al., 2014b) for the statistical computing environment R. However, is not currently supported by standard inventory (forecasting) software packages whereas both top-down and bottom-up approaches are. It is important to note that this interesting approach is operationally very complex but consistent with current practices of many managers and thus might fit very well with the mind-set of many potential users.

3.5. Current gaps

Except for Fliedner and Lawrence (1995) and the literature referred to in sub-section 3.2 (in the context of determining groups for optimal determination of seasonal and other group-based indices), the formation of groups within a hierarchy in order to maximize the accuracy gains from hierarchical forecasting has received very little attention in either the academic or the practitioner literature. Indeed, practitioners usually rely on predefined hierarchies based on non-forecasting considerations, which may or may not make sense for hierarchical forecasting. Given the intuitive appeal of grouping methods for hierarchical forecasting, and the promising empirical results that have been presented in this area, more research into the formation of hierarchical groups would appear to be merited.

The previous subsections addressed hierarchically consistent point forecasts, which will of course be the point of departure for any decisions made in the supply chain. However, no supply chain can function without buffer and safety stocks, which are set based on future expected volatilities of demands, i.e., prediction intervals or density forecasts, which should of course incorporate the covariance structure of the reconciled hierarchical forecasts. Unfortunately, there appears to be very little work investigating density forecasts in an aggregation or supply chain context, except for the very specific cases discussed in sub-section 2.1 and *Appendix C*.

This entire section has for its overarching theme the reconciliation of cross-sectional hierarchical forecasts. As discussed, there has been much research on possible reconciliation techniques from a theoretical and mathematical point of view. However, in practice, often divergent forecasts originate in different business units in a firm or from different nodes in a supply chain, which then need to be

reconciled in a Sales & Operations Planning (S&OP) process (Grimson and Pyke, 2007). Rarely will all participants agree on a single algorithm to reconcile forecasts. More likely, group-based reconciliation will take place, at least for more important sub-hierarchies. In this context, the group dynamics of real-world decision makers reconciling forecasts, e.g., in a Delphi environment, become relevant – and there does not seem to be much research on this topic (de Almeida et al., 2015), in contrast to a very active research community working on single-person cognitive biases in forecasting. Interpersonal, rather than statistical, reconciliation processes were only very recently suggested in the literature (Spithourakis et al., 2015) and more research into their potential effectiveness should be of great value to real world practices. [At this point it is important to note that a recent special issue of the International Journal of Forecasting (2011, volume 27) on group judgmental forecasting does not focus on supply chains.] In particular, the effects of automatic reconciliation (optimal, top-down, bottom-up, middle-out) are yet to be understood when incorporating judgment at various levels of the hierarchy.

4. SUPPLY CHAIN FORECASTING: *TIME*

4.1. *Demand characteristics*

The Auto-Regressive Integrated Moving Average (ARIMA) framework of analysis has been most useful for research in supply chain forecasting; all the developments described in Section 2 (Length) and many of the developments discussed in Section 3 (Depth) (and later on in this current Section, and the relevant Appendices), assume that the underlying demand structure may be represented by an ARIMA form. However, when the demand is intermittent (count data) then Integer ARMA (INARMA) models need to be considered. Such models have had applications in different areas such as medical science and economics but unfortunately have not been extensively considered in a supply chain context.

Both ARIMA and INARMA frameworks insightfully distinguish between stochastic underlying patterns. However, they do not readily lend themselves to practical forecasting applications and it is true to say that neither framework is being widely used in practice for forecasting purposes. Consequently, although these frameworks may facilitate modelling of demand series, they would appear to be less helpful in terms of actual extrapolation needs.

Alternatively, exponential smoothing methods have been given a statistical foundation by the use of state space models with a *single source of error* (SSOE) (for the linear form of which equivalent ARIMA models exist). State space models distinguish between an observation (actual demand) and system / state (underlying structure) equation, each being associated with an error (Gaussian white noise process) term. Whereas the *multiple source of error* (MSOE) formulation of these models assumes distinct error terms (for the observation and system equation) with no cross-correlation, in the SSOE models the same error term appears in both equations. This has been claimed to be preferable because it allows the state space formulation of non-linear as well as linear cases, and, in more practical terms, it also allows the state equations to be expressed in a form which coincides with the error-correction form of the usual smoothing equations. That is, the SSOE approach enables one to see directly the relevance of exponential smoothing methods. Nevertheless, it has been criticised as not being a model in an Operational Research sense (Johnston, 2000), since it works backwards from the *method* to the *model*, rather than the other way around which is what we strive to achieve. (A model should offer the

opportunity to gain insights on real data and then derive properties of various estimators; see also Johnston et al., 1986.)

Although a classification of stochastic processes is most useful for theoretical analysis, a classification of forecasting methods is more relevant to practical forecasting needs. Pegels (1969) classified exponential smoothing methods according to factors such as the inclusion of trend and seasonality. However, the classification did not include methods for intermittent demand (such as Croston's method, for example (Croston, 1972) or the Syntetos-Boylan Approximation (SBA) (Syntetos and Boylan, 2005)) even after extension by Gardner (2006). If we move sufficiently low with regards to the level of granularity of demand data we, unavoidably, end up with intermittence. In addition, in a service parts context demand series are almost invariably intermittent in nature calling for relevant estimators and classification schemes that distinguish between them. In a supply chain context, we need to take into account the fact that demand series are often associated with occasional demand occurrences interspersed with many zero demand observations. Thus, a solution that recognises this very supply chain characteristic need to be adopted.

Johnston and Boylan (1996) and Syntetos et al. (2005) argued that previous attempts appeared in the literature, as well as standard classification schemes employed by software packages, to distinguish between 'fast' and intermittent demand items are ad-hoc both in terms of the criteria employed for classification purposes but even more importantly in terms of the cut-off values assigned to such criteria.

Johnston and Boylan (1996) considered an important question: what is the degree of intermittence that renders Croston's method more appropriate than SES. The researchers worked on the premise that it is preferable to compare directly estimation procedures for the purpose of establishing regions of superior performance and then categorize demand based on the results ^(NOTE 7). They compared SES and Croston based on their simulated Mean Squared Error (MSE) and under a wide range of realistic assumptions as far as the demand generation process was concerned. They came to the conclusion that Croston should be preferred to SES for inter-demand intervals greater than 1.25 forecast revision periods. The contribution of their work lies more in the identification of the inter-demand interval as a classification parameter rather than the exact value (cut-off point) assigned to it.

Syntetos et al. (2005) extended the work discussed above to consider theoretically derived MSE expressions. They compared the theoretical performance of SES, Croston and SBA through their theoretical MSE over a fixed lead time of duration L . Such comparisons resulted in regions where one method performs better than others and enabled the selection of specific methods under specific demand characteristics. All three methods were assumed to be employed with a common smoothing constant α and the analysis was subject to a stationary underlying demand framework and the assumption of a Bernoulli demand arrival process.

The MSE over a fixed lead time of duration L , assuming error auto-correlation, is calculated as follows (Strijbosch et al. 2000, Syntetos et al. 2005):

$$MSE_{LT} = L \{ L \text{Var}(\text{Estimate}) + L \text{Bias}^2 + \text{Var}(\text{Actual Demand}) \} \quad (4)$$

It follows from equation (4) that the MSE over lead time L of Method A is greater than the MSE over lead time L of Method B if and only if

$$\text{Var}(\text{Estimate})_A + \text{Bias}_A^2 > \text{Var}(\text{Estimate})_B + \text{Bias}_B^2 \quad (5)$$

where the subscripts refer to the forecasting methods employed.

Considering inequality (5) it is obvious that the comparison between any two estimation procedures is only in terms of the bias and the variance of their one step-ahead estimates. The comparisons resulted in operationalised classification schemes along the lines presented in Figure 3. In addition to the average inter-demand interval (p), the squared coefficient of variation of the demand sizes (CV^2) was also identified as an important parameter to distinguish between alternative methods' performance. The empirical utility of the schemes discussed above has been confirmed in many studies (e.g., Ghobbar and Friend 2002, Regattieri et al. 2005, Boylan et al. 2008). Further refinements of these schemes (and empirical evidence on their performance) are offered by Kostenko and Hyndman (2006) and Heinecke et al. (2013).

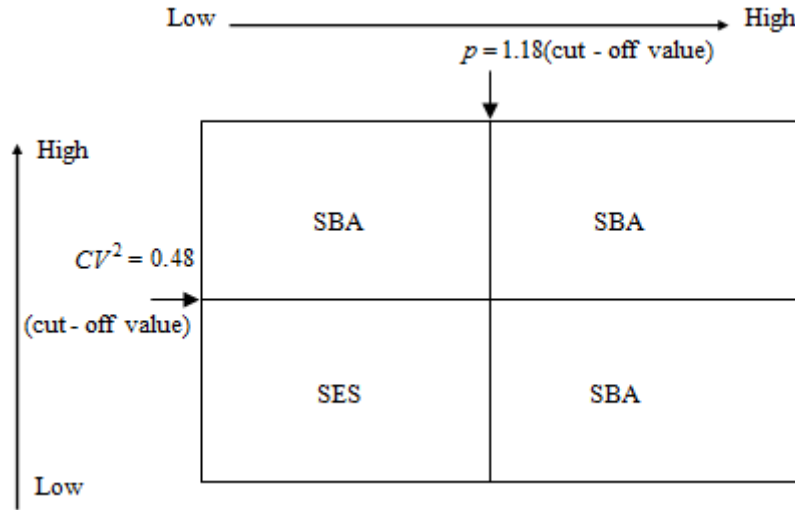


Figure 3. SKU classification based on the comparative performance of SBA and SES (all points in time) for $\alpha = 0.15$

So far, we have juxtaposed intermittent and non-intermittent demand time series. In practice, a given time series may switch between intermittency and non-intermittency. For instance, sales may be intermittent during most of the time, but turn non-intermittent during promotions, during high season (e.g., ice cream or air conditioning spare parts in summer, or heating, headlight or bumper automotive spare parts in winter), or before certain calendar events (e.g., “typical” presents like jewellery, perfume or chocolates before Valentine’s day or Christmas). In such cases, detecting the regimes and appropriate change points between regimes is an important pre-processing step to apply to a time series before selecting the forecasting model. To date, there has been little research on this matter.

An important question that emerges at this point is whether there may be alternative ways to deal with intermittence - other than introducing methods specifically designed for such demand patterns, like Croston or SBA, and attempting to distinguish between them. Such an approach is aggregating demand in lower frequency time units (say from daily to weekly) thereby reducing (or even eliminating) the presence of zero observations. Temporal aggregation is known to generally reduce demand volatility and has been identified as one of the most important areas in a service parts forecasting

context urgently requiring further research (Gardner, 2011). This is discussed in more detail in the following sub-section.

4.2. Temporal aggregation

Forecasting by temporal aggregation is the process of aggregating demands from higher frequency to lower-frequency time buckets - for example aggregating daily data to weekly - and using the aggregate time series to generate forecasts (Syntetos, 2014). This intuitively appealing approach will very often (but not always) reduce demand uncertainty for both fast and slow moving items. However, the benefits of temporal aggregation may not be well-understood by managers and applications are not typically supported by forecasting software – in contrast with cross-sectional aggregation that is supported by inventory and CRM packages, amongst others (please refer also to the previous section).

There are two forms of temporal aggregation: *non-overlapping* and *overlapping*. The first divides the historical information into consecutive non-overlapping blocks of equal length. In overlapping aggregation, the blocks are also of equal lengths but, at each period, the oldest observation is dropped and the newest is included. In this case, the data are actually moving sub-totals of demand history. (Calendar adjustments may sometimes be required when aggregating daily or weekly data into monthly, in which case the blocks will not have the exact same length, e.g. monthly blocks will consist of 28, 29, 30 or 31 days. This may be more of a problem in overlapping rather than non-overlapping aggregation.)

Non-overlapping aggregation results in a considerable reduction of the number of periods; say from 21 daily demands to 3 weekly demands. Practitioners have expressed concerns over the natural loss of information associated with temporal aggregation. However, this concern is most relevant for short demand histories (and non-overlapping aggregation). Should long demand series be available the loss of information resulting from aggregation needs to be contrasted to the potentially improved forecast accuracy associated with the uncertainty reduction.

Temporal aggregation has been extensively used in engineering and physics for signal processing - mainly smoothing for noise reduction and identifying 'true' signals (see, for example, Carstensen et al., 2013); it has also found popularity among scholars in financial time series analysis in the quest of more effective ways to aggregate volatility estimators (Meddahi and Renault, 2004) or time series models in general (Silvestrini and Veredas, 2008). Despite this wide attention from other academic disciplines, *aggregation* has never attracted considerable attention in an Operational Research / Management Science context. Although Willemain et al. (1994) pointed towards that direction, the effects of aggregation were not considered until 2011 when Nikolopoulos et al. empirically assessed the relevant benefits in an intermittent demand forecasting context.

As discussed in the previous sub-section, demand data is frequently *intermittent*. In this case, aggregation of demand into lower-frequency time buckets reduces the presence of zero observations – there are fewer months than weeks with zero sales – which can improve forecast performance. Intermittent demand items (such as spare parts) pose considerable forecasting challenges, both in terms of capturing underlying time series characteristics and fitting standard forecasting models. Temporal aggregation is known to be applied widely in practical military settings (very sparse data), the after-sales industry (service parts) and elsewhere. Nikolopoulos et al. (2011) demonstrated the value of (non-overlapping) temporal aggregation in such a context. Although the paper under concern

was exploratory in nature, it provided some much-needed empirical evidence of the potential benefits of using temporal aggregation in a supply chain context by means of experimenting with the demand histories of 5,000 stock keeping units from the Royal Air Force (UK). In addition to resulting in improved forecast accuracy, the proposed temporal aggregation forecasting framework (referred to as ADIDA: Aggregate Disaggregate Intermittent Demand Approach to forecasting) was also found to lead to stock control performance improvements (Babai et al., 2012) ^(NOTE 8) and to lend itself to application in non-intermittent demand contexts as well.

Spithourakis et al. (2011) provided the first results for fast-moving data by demonstrating that in the M3 competition almost all competing methods could have benefited in terms of forecasting accuracy if applied via the ADIDA temporal aggregation framework. ADIDA has been found effectively to operate as a forecasting method “self-improving” mechanism; changing the data series features through frequency transformation may help extrapolation methods achieve better accuracy performance. Building on those empirical results, some theoretical analysis was presented for data following an AR(1) and MA(1) process by Rostami-Tabar et al. (2013) and a more general ARMA(1,1) process by Rostami-Tabar et al. (2014). This analysis led to theoretically determined optimum aggregation levels for the ADIDA framework (as opposed to the empirically determined optimum levels of aggregation suggested by Nikolopoulos et al., 2011) under specific demand generation processes). Further theoretical developments for both fast and slow moving data were presented by Spithourakis et al. (2014) providing theoretical bounds for the temporal aggregation process to be leading to improvements in forecasting accuracy.

Kourentzes et al. (2014) extended the fundamental idea of ADIDA by means of estimating time series structural components across multiple frequencies. Multiple time series are constructed from the original time series, using temporal aggregation, and automatically selected exponential smoothing methods are used to produce forecasts. Subsequently, the time series forecasts from each aggregation level are combined so as to construct the final forecast. Empirical evaluation of the proposed framework demonstrated significant improvements in forecasting accuracy for longer-term forecast horizons.

Petropoulos and Kourentzes (2015) further extended the research discussed above by means of considering also extrapolation method combinations. They empirically evaluated forecast combinations for intermittent demand data using both method and temporal combinations of forecasts. The first are based on combinations of different methods on the same time series, while the latter use combinations of forecasts produced on different aggregated frequencies of the time series, based on temporal aggregation. Results indicated that appropriate combinations lead to improved forecasting performance over single methods, as well as simplifying the forecasting process by limiting the need for manual selection of methods.

The work described thus far in this sub-section refers to the process of estimating mean demand. However, forecasting of the demand variance is equally important in a supply chain (inventory control) context. The effect of temporal aggregation for variance estimation remains to date a largely unexplored area with only a minimal number of contributions offered in the literature. In particular, Syntetos and Boylan (2006) suggested a cumulative (smoothed) Mean Squared Error (MSE) procedure for forecasting the variance of the lead time demand. The traditional approach to such an estimation procedure is to consider the one-step-ahead (smoothed) MSE and multiply it by the (constant) lead time, assuming independence of the forecast errors over time. However, even in the presence of

stationary independent demand, the forecast errors will be auto-correlated (even though demand is not – please refer also to equation (4)) (Syntetos et al., 2005) and thus the traditional approach will under-estimate variability (and, further, lead to safety stocks lower than those required to meet a prescribed service target) (Syntetos et al., 2015b). Direct updating of the cumulative lead time forecast error avoids this problem, in addition to having an intuitive appeal and being easy to implement in practice. Despite a number of counter arguments to this approach [e.g., i) different forecasting periods may be needed for different SKUs; ii) situations with stochastic lead times cannot be accommodated] this remains the sole proposition in the literature for exploiting the concept of temporal aggregation for variance estimation. At this point it should be mentioned that matching the aggregation level to the (constant) lead time has also been found to perform well for mean demand estimation both in terms of forecast accuracy (Nikolopoulos et al., 2011) and stock control performance (Babai et al., 2012).

4.3. Current gaps

Temporal aggregation appears to be a very promising approach, both for fast and intermittent demand items. However, there is considerable scope for further work in this area both for mean demand and demand variance estimation. The latter is a relatively unexplored area but a most important one for inventory management as safety stocks (inventory investments) link explicitly to the estimated demand variance. Further, the effects of temporal aggregation on the volatility of the series are not clearly understood. Temporal aggregation very often, but *not* always, reduces the coefficient of variation (CV) of demand. Theoretically identifying the cases where temporal aggregation increases volatility should be of great value for practical forecasting applications.

Another interesting line of research would be identifying the extent to which temporal aggregation could help us decide what data collection frequencies and respective forecasting horizons we should be using in practice, as for example switching from observing demand weekly to bimonthly may lead to significant improvements in our forecasting capability. This is very important if we take into account that in many application areas we are interested in a cumulative forecast. This issue has been considered by Nikolopoulos et al. (2011) providing evidence that tying the forecasting horizon to the lead time could lead to considerable forecast accuracy improvement, when measuring performance at that level of aggregation (i.e. equal to the lead time) (see also Nikolopoulos and Petropoulos, 2015, for a discussion of this issue in the context of high uncertainty).

A final important question, in terms of data characterisation, is which demand framework we should be using in practice to distinguish between alternative demand patterns. Both ARIMA and INARMA frameworks insightfully distinguish between stochastic underlying patterns. However, they are associated with estimation procedures that are not well-used in practice. Pegels' classification does cover methods that are being used in practice, but fails to reflect the fact that demand is very often intermittent, and thus estimators specifically designed for such demand patterns need to be considered. In that respect the classification scheme(s) proposed by Syntetos et al. (2005) are relevant, although their construction is limited to MSE considerations and specific forecasting methods, in addition to not taking into account inventory implications. Distinguishing between alternative demand patterns is a fundamental issue in supply chain forecasting and further work in this area would be of great theoretical and practical value.

5. JUDGMENT

5.1 Interactions with the time dimension

Some aspects of the *time* dimension considered in this paper link to the potential usefulness of incorporating judgment into the forecasting process. First, it is clear that the application of statistical forecasting methods requires long demand histories. Should only short demand histories be available statistical methods cannot be implemented (or any implementation would be sub-optimal since limited data precludes the identification of the underlying structure and components of the series). Note that such long demand histories need not necessarily come from the time series to be forecasted itself. Indeed, forecasting for new products is often performed by selecting appropriate “predecessor” or “similar” products to the one to be forecasted ^(NOTE 9). By fitting appropriate models to these longer time series of sales of “similar” products, one can forecast a new fashion product’s entire lifecycle, or a new durable good’s adoption curve (Goodwin et al., 2014). A related approach is not to use similar products, but similar locations, by offering a new product initially at a small sample of test stores and extrapolating to other stores (e.g., Fisher and Rajaram, 2000), which is, however, only possible if the goods already physically exist, so requires separate production runs or can only be used for short- or medium-term forecasts.

Either one of these approaches, of course, leads to the question of how to select “similar” products (locations) that are likely to have sales patterns that match the new product (location) to be forecasted. Approaches include simply using the product category, assessing product/location attributes for explanatory power using machine learning techniques, clustering locations on sales patterns, and various judgmental approaches. In any case, forecasts based solely on similar products will be highly uncertain and should be compared to and updated based on initial sales of the new product as soon as these become available.

For short time series that at least allow *some* statistical model estimation, complementing statistical forecasts with managerial judgment may lead to improved forecast performance. (This is also very important when we have special events, like a product promotion, on the horizon and have limited data to estimate their effects on sales, or when there are changes in prices or assortments.) Such judgmental inputs may take various forms, from a simple override of the forecast parameter values, to overriding model and method selection to directly judgmentally adjusting the statistical forecasts. In this paper only the last case is considered. Although the word ‘short’ needs to be qualified (and the cut-off points between the scenarios discussed above are clearly context-based) the message we wish to convey is that the relevance of judgmental inputs into the forecasting process is very much related to the length of the historical series.

Further, the comparative performance of judgmental and statistical forecasting also relates to the forecast horizon. Statistical forecasting should usually be preferred for very short time horizons. Such a choice is also necessitated if there is a high volume of short-term decisions that preclude the use of judgmental forecasts. For example, a supermarket that requires forecasts every half day in order to replenish stock for say 70,000 SKUs could not rely upon judgmental approaches (unless by exception). At the other extreme, it is clear that the longer the forecast horizon the greater the probability that the series will undergo some structural changes and (explanatory) factors that have not been considered in the past will come into play. In such cases pure judgmental forecasts should perform

better. In contrast to short term forecasting, it is the nature of the decisions involved in long term forecasting as well as the likely change in demand dynamics over the forecasting horizon that calls for human input rather than reliance on a statistical model. It makes sense then that an integrated approach would be more useful for medium forecast horizons where useful information external to the statistical model may be brought by managers into the forecasting process by means of adjusting the statistical forecasts.

This above discussion is graphically depicted in the following figure (Figure 4) where we show the type of forecasting in relation to the demand history and forecast horizon.

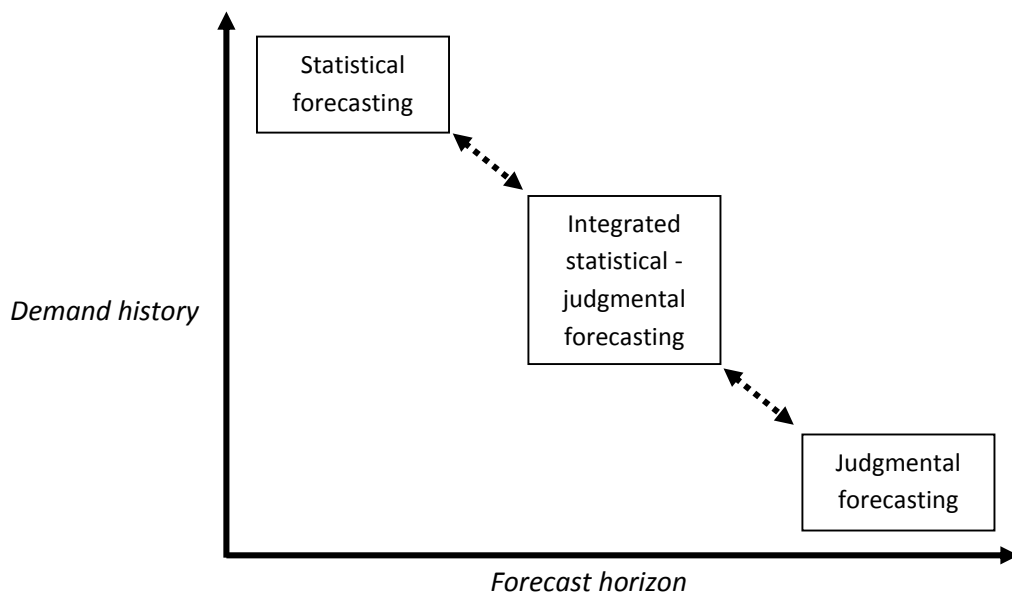


Figure 4. Type of forecasting in relation to the demand history and forecast horizon

5.2 Judgmental adjustments

In this review, given the SCF context, we will discuss the seminal contributions for the middle case described in Figure 4, focussing particularly on judgmental interventions to statistical forecasts (often referred to in the literature as judgmental *adjustments* or judgmental *revisions*). This is because the case of pure judgmental forecasting relates to single echelon tasks and it involves a plethora of studies from the field of psychology dealing with cognitive functions and biases – an area that is clearly outside the scope of this review paper. On the other hand judgmental interventions involve the use of relevant and, hopefully reliable, information coming from the whole supply chain, from shortages in supply to increases in end-product demand and consumer behaviour. Such information is difficult to integrate in theoretical models and empirical evidence has been the predominant source of useful insights that have helped to develop our understanding in this area.

Although judgmentally adjusting statistically derived forecasts is a very common practice in industry (Fildes and Goodwin, 2007), this is an area that has attracted relatively little attention in the academic literature. There are a few empirical studies available analysing the effectiveness of judgmentally adjusted forecasts and their performance in comparison with the base-case (the unadjusted statistical forecasts).

The first large-scale empirical evidence was provided by Mathews and Diamantopoulos with a series of contributions (1986, 1989, 1990, 1992) showing that judgmental revisions improve accuracy. Their results were verified over a sample of more than 900 products from a major UK pharmaceutical manufacturer and retailer. The first study (1986) examined the improvement of judgmental interventions over one quarter and the main result was that the revised forecasts had less variance. The longitudinal extension of this study came three years later (1989) where data and forecasts over six consecutive quarters were examined. Stronger evidence was found of improvements in the forecasting process as a result of the judgmental interventions. The third study (1990) revealed that both the magnitude and direction of the adjustments is affected by both product-specific and reviewer-specific factors, but any potential accuracy improvements can be attributed only to the latter. The final study (1992) demonstrated only marginal differences in the forecasting accuracy related to judgmentally adjusted versus non-adjusted forecasts.

This stream of research was not taken up until about 15 years later by Fildes et al. (2009); data were collected on more than 60,000 forecasts and judgmental adjustments from four supply-chain companies. In three of the companies using monthly data (home cleaning, pharmaceutical and food & beverages) judgmental adjustments were found, on average, to increase forecast accuracy. However, a detailed analysis revealed that, while the relatively larger adjustments tended to lead to greater average improvements in accuracy, smaller adjustments often damaged accuracy. Furthermore, positive adjustments were much less likely to improve accuracy than negative adjustments. They were also made in the wrong direction more frequently, suggesting a bias towards optimism. To take this into account some models were subsequently developed to eradicate relevant biases. In the fourth company (major UK retailer) that was adjusting forecasts at a smaller percentage but observing demand more frequently (weekly data), judgmental adjustments were actually found to decrease forecast accuracy.

The richness of the data gathered through this project led for the first time to the analysis of the effects of integrating management judgment into intermittent demand forecasts (Syntetos et al., 2009) as well as the evaluation of the utility of judgmental adjustments through inventory control (rather than) forecast accuracy metrics (Syntetos et al., 2010). The latter study found an impressive amplification effect: a small improvement in forecasting accuracy via judgmental adjustments may result in far more significant inventory investment savings or increased customer service levels.

At more or less the same time, an equally important database was constructed by Franses and Legerstee containing SKU level statistical and judgmentally adjusted forecasts from the pharmaceutical industry, across thirty-five (35) countries and seven (7) distinct product categories. This led to a series of publications (Franses and Legerstee, 2009, 2010, 2011a, 2011b, 2013) all dealing with different aspects of that database and testing hypotheses similar to those assessed by Fildes et al. (2009). An important conclusion (contradictory to the finding by Fildes et al., 2009) was reached by Franses and Legerstee (2010) that judgmentally adjusted forecasts are, at best, equally good to pure statistical forecasts, but most often are worse, with the profound deficiency being attributed to experts' tendency to put too much weight on their own contribution. More recently Petropoulos et al. (2015) looked at the same database examining some behavioural aspects and provided evidence that experts change their behaviour when performing judgmental adjustments to statistical forecasts after 'big losses' (judgmental adjustments that significantly decrease forecast accuracy).

Acknowledging the small number of empirical studies in the field, the almost complete absence of formal OR modelling in order to offer insights into the impact of such judgmental adjustments on supply chain performance and the potential mitigation mechanisms necessary are even more surprising. We could identify only one study by Syntetos et al. (2011) that employed a System Dynamics modelling approach to theoretically evaluate the effects of forecasting and order replenishment adjustments for a wide range of scenarios, involving three different inventory policies, seven different (combinations of) the supply chain points of intervention and four different (combinations of) types of judgmental intervention (optimistic and pessimistic).

5.3. Current gaps

It is evident, given the small number of empirical and theoretical studies as well as the inconclusive empirical results, that there is still considerable space for empirical investigations and more importantly theoretical and methodological contributions towards identifying the specific conditions under which judgmental adjustments do lead to improvements in forecast accuracy and forecast utility. Another promising direction for further research is the assessment of the implications of superimposing judgmental adjustments on other layers of the supply chain and/or the decision making spectrum and the resulting interactions between such judgmental interventions. (An example is the very recent investigation of the effects of integrating management judgment into stock replenishment decisions by Syntetos et al., 2015a).

There is also little evidence on the effectiveness of judgmental interventions in relation to the rationale behind such interventions. There is not even a single study to date that links the adjustments to some justification behind these adjustments ^(NOTE 10). Further, our understanding on the effects of successive judgmental interventions is very limited. How do people adjust forecasts when they know that these forecasts have already been subject to modification by someone else in the supply chain? How do judgmentally adjusted forecasts interact with judgmental adjustments further down the decision making process? For example, practitioners confirm that replenishment orders, transportation routes, production schedules and so on, all of which rely upon a (possibly judgmentally adjusted) forecast, are frequently subject to additional judgmental modifications.

6. DATA AND SOFTWARE

As described in the previous sections, there have been major advances on the algorithmic and statistical sides of the problem of forecasting in a supply chain context. Nevertheless, major problems remain from a software and especially from a data perspective.

Academic research has in recent years devoted more and more attention to investigating desirable feature qualities of Forecasting Support Systems (FSS; Fildes and Goodwin, 2013). Apart from general User Experience (UX)/User Interface (UI) requirements, specific FSS features need to take known cognitive and behavioural biases in (frequently non-expert) forecasters into account (Fildes et al., 2009) and should include, e.g., displaying statistical and business information about a time series to be forecast, allowing for judgmental adjustments, providing for or even requiring the user to give reasons for adjustments, tracking, monitoring and evaluating adjustments, and enabling users to share information within the system (Spithourakis et al., 2015). Unfortunately, we know of no commercially

available FSS in this sense. In our experience, most forecasters employ a mixture of tools, using Microsoft Excel or a dedicated forecasting tool for forecasting, capturing forecasts again in Excel or in the company ERP and overwriting forecasts based on judgment again in Excel, with no provision for tracking reasons for adjustment or evaluating whether adjustments in fact improved the forecast.

There is a gap to be bridged between statistical (and behavioural) theory and operational and economic practice. Software developers play a key role in addressing this gap. However, it must be kept in mind that developers usually receive just as little training in statistics or forecasting as statisticians and forecasters receive in software development (cf. the well-known statistical/numerical weaknesses in Microsoft Excel, McCullough and Heiser, 2008). Modern software development is a vast, complex and rapidly changing field – just as statistics. It makes little sense to train software developers in rudimentary statistics, nor to train statisticians in rudimentary software development practices, to have them write forecasting software that will, necessarily, perform rudimentarily. Instead, what is needed is a close cross-functional collaboration between software developers and statisticians/forecasters, all of which speak each others' languages to an extent that allows them to create working forecasting software or full FSS. Such a close collaboration will be easier in developing standard software, where economies of scale allow retaining statistical specialists, than in custom development, where statistical knowledge can, at best, be included only on an ad hoc basis for cost reasons.

The recent trend towards “Big Data”, which is facilitated by new developments in database technology (Januschowski et al., 2013), has major implications for forecasting in supply chains. For instance, data are kept at finer and finer granularities – whereas in the past, retailers would store aggregated sales on SKU \times store \times week basis, today they commonly keep far more disaggregated data, at least on a daily level. Increasingly, data are also stored disaggregated per offer (“on day t , we sold 8 units at regular price, 6 units on a buy-one-get-one offer available with a newspaper coupon, and 14 units at 30% off available to shopper card holders”), or even on transaction log (TLOG) or basket level data. Highly granular data raise the question of what aggregation level to forecast on, as detailed in previous sections, and frequently lead to the problem of forecasting intermittent time series.

A related dynamic is the increased availability of information linking specific purchases to specific customers. This has of course always been possible for mail order and online retailers (where online retailers also know the customer's browsing history on their site, or even on other sites via cookies, while mail order retailers are of course limited to knowing what customers actually ordered in the end), but the increasing ubiquity of shopper cards makes this information available to an extent to traditional brick-and-mortar retailers, as well. Retailers have mainly used this information in data mining and association analyses to produce personalised recommendations and discount coupons (Humby et al., 2008), but it may be possible to also leverage it in forecasting.

At the same time, established retailers face increased competition from new entrants into the retail market, be it pure online players or new entrants into a geographical market, like German discounters taking market share from the leading UK grocer Tesco. In addition, consumer products manufacturers, especially in the fashion sector, increasingly integrate vertically in the supply chain by opening their own web or brick-and-mortar stores, competing directly with established retailers. Incumbents frequently react by increasing the scope and complexity of promotions they offer, which in turn increases the complexity of the data they have and of the forecasting task they face (see above).

In addition, participants in the supply chain increasingly aim at leveraging social media information in forecasts. Positive or negative publicity can have a major impact on customers' demand, but including this information in forecasts is so far an unsolved problem at the intersection of forecasting algorithmics, textual analysis and data management.

The common theme of these developments is that more and more data of higher and higher complexity is in principle available to be leveraged in supply chain forecasting. However, established Enterprise Resource Planning (ERP) systems typically spread these data over many disparate tables and/or separate systems. One apparently banal but nevertheless important challenge for software providers thus is to access and combine all relevant data – including the use of APIs and feeds for external data, e.g., from social media sites – and to ensure data consistency. This is not a forecasting issue properly speaking, but a data management problem which needs to be solved before causal forecasting methods can even be brought to bear.

A related problem lies in the quality of the data available. For instance, DeHoratius and Raman (2008) found that 65% out of nearly 370,000 inventory records examined at a retailer's 37 stores were inaccurate. Possible reasons for this level of data inaccuracy include inexact deliveries, undocumented goods movements between the backroom and the shop floor, theft, untracked spoilage and customers or store associates shifting product between locations in the store. This particular data quality problem has two consequences in the present context of forecasting in supply chains: first, forecasts will be biased low by undetected out-of-stocks. And second, even good forecasts will lead to incorrect reordering decisions and suboptimal inventories if paired with incorrect system inventories. Other forecast-relevant data issues include miscoded promotions, incorrect prices, or untracked special influences like unscheduled store closings or employee strikes. Thus, data quality problems limit the impact of improved forecasting algorithms or processes such as information sharing on improving inventory positions, and data quality is one key issue that needs to be addressed by retailers.

In this section, we have focused on the retailer as the supply chain node which is closest to the end customer. Of course, these dynamics propagate to higher levels in the supply chain via the information and materials flows (Figure 1), and also by consumer product manufacturers expanding into the retail sector. Thus, the issues identified above also apply to higher levels in the supply chain.

Overall, the move towards more and more data counteracts the trend towards faster and faster implementation of analytics and forecasting software (Moore's law): retailers' expectations outstrip software vendors' capabilities (Parkinson's law). Consequently, the quest for improved performance of databases and algorithm implementations is not likely to end any time soon.

7. GAPS, CHALLENGES AND THE FUTURE

7.1. Where do the gaps come from?

The paper has discussed thus far a number of gaps between forecasting theory and practice. Academics are often accused of not researching issues of practical relevance, resulting in forecasting theory lagging behind real world practices. Although this is often true, the converse is also frequently the case. Practitioners and software packages do not often employ methods or approaches that have

been shown (in academic studies) to be both effective and easy to implement, thus leading to real world practices lagging behind theoretical developments.

Failed academic attempts to serve real world applications may stem from the fact that the issues being tackled are simply of no practical importance at all. In addition, though, the extremely fast developments and changes in the area of supply chain forecasting render relevant theoretical developments a very difficult exercise as the ‘problem setting’ keeps on evolving at a very high speed. For example, the introduction of multichannel operations discussed in the previous section introduce a very challenging interface between the dimension of *length* and *depth*. Indeed, our own theoretical framework may be subject to modifications as shown below (please see Figure 5).

Traditional upstream propagation of demand does not reflect the fact that consumers may in fact order directly at levels further up than the retailer. Consider for example the case that the consumers may order directly at the wholesaling or manufacturing level that also faces demand as propagated through the supply chain. This would naturally generate the need to look at a different form of aggregation; one that combines (sums) the propagated demand up to a particular supply chain level and the direct orders received from consumers at that level. Exactly the same concept applies (perhaps even more often) to orders placed by the retailers directly to the manufacturers.

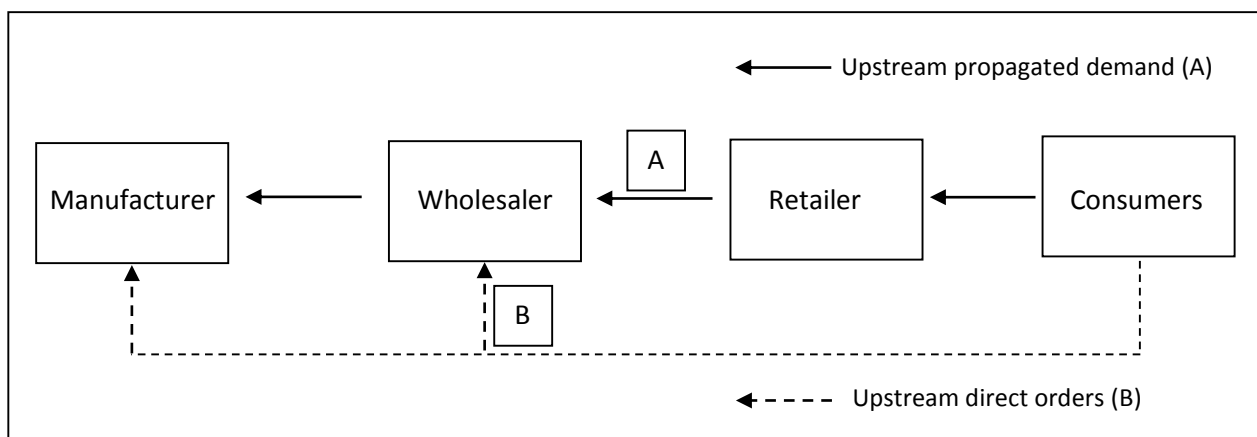


Figure 5. Aggregation of upstream propagated demand (A) and direct orders (B): $A+B$

However, and as discussed above, practitioners (and software manufacturers) also contribute to the gap between theory and practice by not employing methods and approaches that would make a considerable difference in their operations. Some indicative examples are the following:

- The classification scheme(s) developed by Syntetos et al. (2005) have been shown in a number of studies to offer considerable benefits in terms of forecasting and stock control and they have indeed been successfully implemented in practice, unfortunately though in very few cases (e.g. by *JDA software*). Similarly, and although not explicitly considered in this paper, the SBA estimator has repeatedly been shown to offer considerable forecast accuracy improvements and to be a very robust estimator, and yet it is being used in only a few real world applications we are aware of (e.g. *Syncron*).
- The literature on judgmental adjustments of statistical forecasts has shown that (large) negative adjustments tend to improve forecast accuracy, whereas (large) positive adjustments result in an opposite effect. Judgmental adjustments of statistical forecasts take place routinely in industry, and yet companies are not introducing authorisation mechanisms

according to which negative adjustments can be freely conducted but positive adjustments require explicit justification or authorisation.

- Although further research is needed in order to fully appreciate the benefits (and drawbacks) of temporal aggregation it is clear that at the very minimum it is worthwhile to experiment with its implications for decision making. Yet, software packages do not support temporal aggregation.
- Cross sectional aggregation for seasonal (trended or untrended) forecasting is both intuitively appealing and has been shown to offer considerable benefits when demand histories are short. However, not many companies utilise such an approach to estimating seasonality.

Before we close this section we should mention that the emphasis of our paper has been on classical time series approaches; causal methods have not been covered. With the recent ‘hype’ around Big Data, many practitioners have been trying to leverage causal effects such as weather forecasts and social media activity. Forecasting solutions are increasingly expected to use Twitter, Facebook and other external leading indicators leading to what may be called ‘predictive analytics’. Although there doesn’t seem to be a clear-cut separation between ‘forecasting’ and ‘predictive analytics’ we take the view that the latter does not usually utilise time series by themselves. Further, identifying the extent to which explanatory variables may indeed help the forecasting task has not received considerable attention in the supply chain context and arguably there has not been much written on the state-of-the-art of using such (relevant) explanatory variables. There is a developing literature in this area in the field of Marketing Analytics but such developments have not been matched in the supply chain discipline. Integration of Marketing and SCM related work to comprehensively establish ‘what can be actually done’ in the field of analytics, coupled with the industry progressing beyond this current ‘hype’ phase to determine ‘what needs to be done’ (other than unrealistically expecting everything to be linked to everything) should be of great value towards further bridging the gap between forecasting theory and practice.

7.2. Current challenges for theorists, practitioners and software developers

In this sub-section we suggest a number of important issues where there is a call for further experimentation and theoretical research.

How to make effective use of information contained in social media? Getting extremely noisy external causal factors right (particularly social media) currently presents practitioners and software developers with a number of challenges. In addition, unifying different data sources, and then making sense of them seems also to be an area requiring further attention.

How to forecast discrete predictive densities? Most theoretical results, as well as most forecasting software packages, make heavy use of the normal distribution. This is defensible in high count time series, in which actual count data can reasonably be approximated by normal distributions. However, exploding data storage and analysis capacities (“Big Data”) are driving a trend towards storing and forecasting data in finer and finer levels of granularity, so that the case for the normal approximation becomes weaker and weaker. To appropriately set safety stocks for highly granular count data, we need to understand *discrete* predictive distributions. There has been remarkably little work in this regard, and research in this direction is urgently needed.

How to structure hierarchies? In any forecasting process, the first step is the identification of the outputs to be forecast. When we require a demand forecast, we must specify the aggregation level needed. In turn, this requires identifying the:

- **Product:** e.g. SKU, item, family, etc.
- **Market / Location:** customer, POS, region, country, business area, etc.
- **Time:** hour, day, week, month, year, etc.

The forecasts needed to support the decision-making process may be aggregate in one or some of those dimensions and granular in others. For example, we may require product forecasts by SKU but market forecasts only by region. How to structure the hierarchy to best forecast this type of hybrid decision-making problem is a topic about which much remains to be learned.

Which are the current forecasting software limitations? A recent collection of articles in a Forecasting Guide published by *Foresight: The International Journal of Applied Forecasting* comes to the clear conclusion that software limitations are impeding progress in hierarchical forecasting. From an inventory perspective, aggregating requirements across a group of products ordered from the same supplier is a very natural thing to do, and such a process is supported by inventory software packages. Similarly, the aggregation of forthcoming requirements from a specific customer or the aggregation of demand in a specific region is highly relevant to marketing and sales initiatives and is supported by customer relationship management (CRM) packages. However, and as discussed in this paper, this is not the case for temporal aggregation. Nor is there support for non-traditional reconciliation approaches, those that rely on combining forecasts. Expanded functionality here should not require considerable interventions. Indeed, sub-section 4.2 discusses a multiple aggregation prediction algorithm (MAPA, Kourentzes et al., 2014; Petropoulos and Kourentzes, 2014) that has been made available in the open-source statistical software package R and provides a template for commercial-software implementation.

How to forecast lead-time demand variance? There is an important issue in inventory forecasting that has not yet been sufficiently addressed. Here, we need forecasts not only of mean demand, but variances as well (to set safety stocks). For instance, even if we were to restrict ourselves to a temporal hierarchy, which of the following is the better route to obtain an estimate of the variance of lead-time demand: i) to forecast the demand variances for each of the next months and add these over the lead time, or ii) to extrapolate directly the cumulative lead-time demand variance (assuming constant lead times)? No clear answer is available here because the issue has not been adequately studied. We need to extend the experiments with temporal aggregation to cover demand-variance estimation.

A related area of neglect is the potential for aggregation techniques to provide superior estimates of quantile or density forecasts and other parameters of forecast distributions in the context of hierarchies. Quantile forecasts, prediction intervals and density forecasts have proven very difficult to obtain for single time series. Aggregation may help or hinder us here, but there has been remarkably little work to show how.

How to combine temporal and cross-sectional hierarchies? Our review has addressed both cross-sectional and temporal hierarchies, but as separable structures. In reality, companies must make decisions in both dimensions. A *unified approach* – one that simultaneously considers choices of

aggregation levels and frequency along multiple dimensions – would be a valuable step in the right direction. The problem with the separation of the cross-sectional and temporal dimensions is that the *right level* of cross-sectional aggregation may vary across time frequencies and vice versa. Procedures that combine forecasts for a cross-sectional hierarchy, such as the optimal reconciliation technique discussed in sub-section 3.4, and procedures that combine forecasts over time frequencies, such as the multiple temporal-aggregation technique discussed in sub-section 4.2, may conceivably be *pooled* to form a holistic – or, idiomatically speaking, *whole-istic* – strategy for forecasting hierarchies.

How to reconcile to achieve consensus? Instead of using statistical procedures to reconcile forecasts across a hierarchy, users can choose to create an *interpersonal reconciliation process*. As proposed in a recent study by Spithourakis et al. (2015), the process may nurture consensus among forecasters and make managers at different levels of a hierarchy interdependent. This too is an area where new research is merited. (Some interesting issues in this area are discussed by Fisher et al., 1994).

How to forecast supply? This paper has focused on forecasting *demand*, but there are also conditions under which we need to forecast *supply* (as well as possibly other variables such as *stock* actually available in the store or in the warehouse, that is affected by errors as shown by DeHoratius and Raman, 2008). For instance, suppliers may not be perfectly dependable, delivering either too late or less than was ordered. In such a case, we need safety stocks not only to buffer against demand volatility, but also against supply volatility. Similarly, many companies operate so-called reverse logistics; that is, they acquire used materials from later stages in the supply chain and remanufacture them. Examples encompass beverage companies receiving empty bottles in deposit schemes, or recyclable materials being collected. In either case, supply is not deterministic and needs to be forecasted. There is a lack of research on specific challenges in this area, in particular on optimal inventory policies in the presence of uncertainty in both supply and demand.

Forecasting Support Systems (FSS). As noted above, there appear to be no commercial FSSs worthy of the name available for productive forecasting. On the one hand, this seems to be an opportunity for software vendors. On the other hand, this very lack calls for an explanation, and market research as to why exactly none of the major forecasting software providers has stepped up to provide a fully-fledged FSS would be an interesting topic for academic researchers.

7.3. Conclusion

Supply Chain Forecasting (SCF) goes beyond the operational task of extrapolating demand requirements at one echelon. Working on this premise, we have critically reviewed the literature attempting to highlight approaches the development of which has its roots in the context of the entire supply chain, rather than a single echelon. A conceptual case was made for four key supply chain dimensions around which the literature was reviewed:

- *Echelon dimension*: leading to certain properties of the demand as it is propagated through the various supply chain levels;
- *Location dimension*: leading to issues of how to cluster locations at different echelons (including customers);
- *Product dimension*: leading to issues of addressing cross-sectional product hierarchies;
- *Time dimension*: related issues that form a different type of a hierarchy (temporal).

Other forms of aggregation are also possible and maybe beneficial in certain decision contexts. For example we may aggregate by customer size or by segment of customers (Boylan and Syntetos, 2015). This forms a natural avenue for further research.

The gaps between theory and practice were identified for each of those four key dimensions followed by a discussion on where these gaps might originate from. The discussion was complemented by reviews of judgment, data and software related constraints and opportunities for further work. Finally, a number of challenges were highlighted to suggest promising and relevant avenues for further research.

The field of SCF is one that evolves very rapidly, offering ample opportunities for relevant academic research and the application of innovative thinking towards solving real world problems. We hope that this review paper will motivate interesting developments in the field and facilitate a more constructive exchange of information and collaboration between academia and industry.

ACKNOWLEDGEMENTS

We would like to thank all our colleagues who contributed to the development of this paper by filling in the relevant questionnaire. We also thank Steve Disney and Fotios Petropoulos at Cardiff University, for their constructive comments on the material presented in Sections 2 and 3 respectively, and Len Tashman, the Editor of *Foresight*, for allowing us to reuse material from the "Hierarchical Forecasting" Guidebook for sub-section 6.2. Finally, we thank the anonymous referees for their constructive comments that greatly helped to improve the content of our paper and its presentation.

NOTES

NOTE 1 The characterization of the information and materials flow upstream and downstream the supply chain respectively serves only discussion purposes and indeed many researchers opt for the presentation of these flows in reverse (vertical) order, or in a horizontal fashion.

NOTE 2 For a comprehensive review of forecasting developments in the field of Operational Research, interested readers are referred to the work of Fildes et al. (2008). Syntetos et al. (2009) provide a review of inventory forecasting methods in the preceding 50 years, and Boylan and Syntetos (2010) review forecasting methods in an intermittent demand (spare parts) context.

NOTE 3 The effects of various forecasting methods on the magnitude of the Bullwhip effect were recently considered by Najafi and Zanjirani Farahani (2014).

NOTE 4 As discussed in the main body of the paper, we consider locations along with customers. A retailer may be one single customer for a CPG manufacturer, but each one of the retailer's distribution centers may need a separate forecast. DCs may differ because of regional particularities or because of operational conditions (e.g., only a few DCs could have the equipment to deal with frozen goods, or a few DCs may be dedicated to deal with promotional quantities).

NOTE 5 Gross and Sohl (1990) found that two simple disaggregations may be particularly effective: i) For each period, calculate the proportion that an individual series contributes towards the aggregate series. Then average the proportions over all periods; ii) Calculate the total, over all periods, for the individual series, and then do the same for the aggregate series. Then calculate the ratio of the former to the latter. Athanasopoulos et al. (2009) refer to the first method as the *average historical proportions* and the second as *the proportions of the historical averages*. For the first approach, they point out that we need not restrict our methods historical proportions. Forecasted proportions can be used instead. Boylan (2010) noted that although this has not been tested on supply chain data, the use of forecasted rather than historical proportions appears to be promising and is a feature typically offered in forecasting software.

NOTE 6 Zotteri et al. (2014) argue that the best level of aggregation for forecasting should be chosen by the forecasters in consideration of the trade-off between sampling error (data inadequate to generate reliable forecasts) and specification error (data too aggregated to represent diverse demand).

NOTE 7 SKU classification for (inventory) forecasting purposes typically works in the opposite way: ad hoc classification rules are used to separate SKUs into categories, followed by the specification of a forecasting method for each of the categories. But if the purpose of classification is the selection of forecasting methods, then it makes more sense to compare directly possible estimators and then categorize demand based on regions of superior forecast performance.

NOTE 8 The fact that a particular forecasting method or approach may perform better than one other in terms of forecast accuracy does not necessarily imply that such benefits carry also over to the implications (utility) of those methods for the decision making process they serve. Put simply, in an inventory control context for example, the fact that estimator X performs better than estimator Y in terms of forecast accuracy (however this is measured) does not mean that estimator X leads to a better trade-off between customer service levels and inventory investment, which is what matters from a practitioner perspective. As such it has become apparent that we need to distinguish between accuracy and accuracy-implication (utility) metrics (Boylan and Syntetos, 2006; Syntetos et al., 2010).

NOTE 9 Related to this is the application of the hierarchical Bayes approach for short time series demonstrated by Yelland (2010) in the context of forecasting the demand for parts at Sun Microsystems. This works by pooling information across products where there is little or no time series history.

NOTE 10 There are also very few studies on behavioural models that may adjust forecasts for weather conditions (short term weather estimates) (e.g. Nikolopoulos and Fildes, 2013) which is a very important issue for many manufacturers.

REFERENCES

Ali, M.M. (2008). *Centralised demand information sharing in supply chains*, Unpublished PhD thesis, Buckinghamshire New University, Brunel University, UK.

Ali, M., and Boylan, J.E. (2010). The value of forecast information sharing in the supply chain. *Foresight*, 18, 14-18.

- Ali, M.M., and Boylan, J.E. (2011). Feasibility principles for downstream demand inference in supply chains. *Journal of the Operational Research Society*, 62, 474-482.
- Ali, M.M., and Boylan, J.E. (2012). On the effect of non-optimal forecasting methods on supply chain downstream demand. *IMA Journal of Management Mathematics*, 23, 81-98.
- Ali, M.M., Boylan, J.E., and Syntetos A.A. (2012). Forecast errors and inventory performance under forecast information sharing. *International Journal of Forecasting*, 28, 830-841.
- Alwan, L.C., Liu, J.J., and Yao, D.-Q. (2003). Stochastic characterization of upstream demand processes in a supply chain. *IIE Transactions*, 35, 207-219.
- Amemiya, T., and Wu, R.Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, 67, 628-632.
- Anderson, O.D. (1975). On a lemma associated with Box, Jenkins and Granger. *Journal of Econometrics*, 3, 151-156.
- Athanasopoulos, G., Ahmed, R.A., and Hyndman, R.J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, 25, 146-166.
- Babai, M.Z., Ali, M.M., and Nikolopoulos, K. (2012). Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *OMEGA: The International Journal of Management Science*, 40, 713-721.
- Barnea, A., and Lakonishok, J. (1980). An analysis of the usefulness of disaggregated accounting data for forecasts of corporate performance. *Decision Sciences*, 11, 17-26.
- Box, G. E. P., and Jenkins, G. M. (1970). *Time series analysis*, Holden-Day, San Francisco.
- Boylan, J.E. (2010). Choosing levels of aggregation for supply chain forecasts. *Foresight*, 18, 9-13.
- Boylan, J.E., Chen, H., Mohammadipour, M, and Syntetos, A.A. (2014). Formation of seasonal groups and application of seasonal indices. *Journal of the Operational Research Society*, 65, 227-241.
- Boylan, J.E., and Syntetos, A.A. (2006). Accuracy and accuracy-implication metrics for intermittent demand. *Foresight*, 4, 39-42.
- Boylan, J.E., and Syntetos, A.A. (2010). Spare parts management: a review of forecasting research and extensions. *IMA Journal of Management Mathematics*, 21, 227-237.
- Boylan, J.E., and Syntetos, A.A. (2015). Supply chain forecasting: the customer dimension. Paper presented at the 27th European Conference on Operational Research (EURO 2015), July 12-15, Glasgow, UK.
- Boylan, J.E., Syntetos, A.A., and Karakostas, G.C. (2008). Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society*, 59, 473-481.
- Brännäs K., Hellstrom, J., and Nordstrom, J. (2002). A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting*, 18, 19-30.

- Caniato, F., Kalchschmidt, M., Ronchi, S., Verganti, R. and Zotteri, G. (2005). Clustering customers to forecast demand. *Production Planning and Control*, 16, 32-43.
- Carstensen, J., Telford, R.J., and Birks, H.J.H. (2013). Diatom flickering prior to regime shift. *Nature*, 498, E11–E12.
- Chen, H., and Boylan, J. E. (2007). Use of individual and group seasonal indexes in subaggregate demand Forecasting. *Journal of the Operational Research Society*, 58, 1660-1671.
- Chen, H., and Boylan, J. E. (2008). Empirical evidence on individual, group and shrinkage seasonal indexes. *International Journal of Forecasting*, 24, 525-534.
- Cheng, T.C.E., and Wu, Y.N. (2005). The impact of information sharing in a two-level supply chain with multiple retailers. *Journal of the Operational Research Society*, 56, 1159-1165.
- Chopra, S., and Meindl, P. (2012). Supply chain management: strategy, planning and operation, 5th edition, Pearson, New Jersey.
- Croston, J.D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23, 289-303.
- Dalhart, G. (1974). Class seasonality – a new approach. *American Production and Inventory Control Society Conference Proceedings*.
- Dangerfield, B.J., and Morris, J.S. (1992). Top-down or bottom-up: aggregate versus disaggregate extrapolations. *International Journal of Forecasting*, 8, 233-241.
- Dawson, P. (2013). Managing Director of DBO Services (<http://www.dboservices.com>). Private communications to the corresponding author.
- de Almeida, M. M. K., Marins, F. A. S., Salgado, A. M. P., Santos, F. C. A., and da Silva, S. L. (2015). Mitigation of the bullwhip effect considering trust and collaboration in supply chain management: a literature review. *The International Journal of Advanced Manufacturing Technology*, 77, 495-513.
- DeHoratius, N., and Raman, A. (2008). Inventory record inaccuracy: an empirical analysis. *Management Science*, 54, 627-641.
- Dekker, M., van Donselaar, K., and Ouwehand, P. (2004). How to use aggregation and combined forecasting to improve seasonal demand forecasts. *International Journal of Production Economics*, 90, 151-167.
- Disney, S., Holmstrom, J., Kaipia, R., and Towill, D.R. (2001). Implementation of a VMI production and inventory control system. Paper presented at the 6th *International Symposium of Logistics*, July 8-10, Salzburg, Austria.
- Fildes, R., and Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37, 570-576.
- Fildes, R., and Goodwin, P. (2013). Forecasting support systems: what we know, what we need to know. *International Journal of Forecasting*, 29, 290-294.

- Fildes, R., Goodwin, P., Lawrence, M., and Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25, 3-23.
- Fildes, R., Nikolopoulos, K., Crone, S., and Syntetos, A.A. (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society*, 59, 1150-1172.
- Fisher, M., Hammond, J.H., Obermeyer, W., and Raman, A. (1994). Making supply meet demand in an uncertain world. *Harvard Business Review*, 72, 83-92.
- Fisher, M., and Rajaram, K. (2000). Accurate retail testing of fashion merchandise: methodology and application. *Marketing Science*, 19, 266-278.
- Fliedner, G. (1999). An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers and Operations Research*, 26, 1133-1149.
- Fliedner, E.B., and Lawrence, B. (1995). Forecasting system parent group formation: an empirical application of cluster analysis. *Journal of Operations Management*, 12, 119-130.
- Franses, P.H., and Legerstee, R. (2009). Properties of expert adjustments on model-based SKU-level forecasts. *International Journal of Forecasting*, 25, 35-47.
- Franses, P.H., and Legerstee, R. (2010). Do experts' adjustments on model-based SKU-level forecasts improve forecast quality? *Journal of Forecasting*, 29, 331-340.
- Franses, P.H., and Legerstee, R. (2011a). Combining SKU-level sales forecasts from models and experts. *Expert Systems with Applications*, 38, 2365-2370.
- Franses, P.H., and Legerstee, R. (2011b). Experts' adjustment to model-based SKU-level forecasts: does the forecast horizon matter? *Journal of the Operational Research Society*, 62, 537-543.
- Franses, P.H., and Legerstee, R. (2013). Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? *International Journal of Forecasting*, 29, 80-87.
- Gardner, E.S., Jr. (2006). Exponential smoothing: the state of the art – part II. *International Journal of Forecasting*, 22, 637-666.
- Gardner, E.S., Jr. (2011). Forecasting for operations. Keynote paper presented at the 31st *International Symposium on Forecasting*, June 27-29, Prague, Czech Republic.
- Ghobbar, A.A., and Friend, C.H. (2002). Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, 8, 221-231.
- Gilbert, K. (2005). An ARIMA supply chain model. *Management Science*, 51, 305-310.
- Goodwin, P., Meeran, S., and Dyussekeneva, K. (2014). The challenges of pre-launch forecasting of adoption time series for new durable products. *International Journal of Forecasting*, 30, 1082-1097.
- Gordon, T., Dangerfield, B., and Morris, J. (1997). Top-down or bottom-up: which is the best approach to forecasting? *The Journal of Business Forecasting*, 16, 13-16.

- Granger, C. W. J., and Morris, M. J. (1976). Time series modeling and interpretation. *Journal of the Royal Statistical Society, Series A*, 139, 246-257.
- Grimson, J. A., and Pyke, D. F. (2007). Sales and operations planning: an exploratory study and framework. *International Journal of Logistics Management*, 18, 322-346.
- Gross, C. W., and Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, 9, 233-254.
- Hammond, J.H. (1994). Barilla Spa (A), Harvard Business School Case 694-046, Harvard University, USA.
- Harvey, A.C. (1993). *Time series models* (Vol. 2), Harvester Wheatsheaf, New York.
- Heinecke, G., Syntetos, A.A., and Wang, W. (2013) Forecasting-based SKU classification. *International Journal of Production Economics*, 143, 455 - 462.
- Holweg, M., Disney, S., Holmstrom, J., and Smaros, J. (2005). Supply chain collaboration: making sense of the strategy continuum. *European Management Journal*, 23, 170-181.
- Hosoda, T., Naim, M.M., Disney, S.M., and Potter, A. (2008). Is there a benefit to sharing market sales information? Linking theory and practice. *Computers and Industrial Engineering*, 54, 315-326.
- Humby, C., Hunt, T., and Phillips, T. (2008). *Scoring points: how Tesco continues to win customer loyalty*. Kogan Page, 2nd rev. ed.
- Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G., and Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, 55, 2579–2589.
- Hyndman, R., and Athanasopoulos, G. (2014). Optimally reconciling forecasts in a hierarchy. *Foresight*, 35, 42-48.
- Hyndman, R.J., and Kostenko, A.V. (2007). Minimum sample size requirements for seasonal forecasting models, *Foresight*, 6, 12-15.
- Hyndman, R.J., Lee, A.J., and Wang, E. (2014a). Fast computation of reconciled forecasts for hierarchical and grouped time series. Working paper 17/14, Department of Econometrics & Business Statistics, Monash University, Australia.
- Hyndman, R.J., Lee, A.J., and Wang, E. (2014b). hts: Hierarchical and grouped time series. URL: cran.r-project.org/package=hts.
- Januschowski, T., Kolassa, S., Lorenz, M., Schwarz, C. (2013). Forecasting with in-memory technology. *Foresight*, 31, 14-20.
- Johnston, F.R. (2000). Lead time demand adjustment or when a model is not a model. *Journal of the Operational Research Society*, 51, 1107-1110.
- Johnston, F.R., and Boylan, J.E. (1996). Forecasting for items with intermittent demand. *Journal of the Operational Research Society*, 47, 113-121.

Johnston, F.R., Harrison, P.J., Marshall, A.S., and France, K.M. (1986). Modelling and the estimation of changing relationships. *The Statistician*, 35, 229-235.

Kostenko, A.V., and Hyndman, R.J. (2006). A note on the categorization of demand patterns. *Journal of the Operational Research Society*, 57, 1256-1257.

Kourentzes, N., Petropoulos, F., and Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30, 291-302.

Lee, H.L., Padmanabhan, V., and Whang, S. (1997). Information distortion in a supply chain: the bullwhip effect. *Management Science*, 43, 546-558.

Lee, H.L., So, K.C., and Tang, C.S. (2000). The value of information sharing in a two-level supply chain. *Management Science*, 46, 626-643.

Luiz, K.H., Pedro, A.M., and Pedro, L.V.P. (1992). The effect of overlapping aggregation on time series models: an application to the unemployment rate in Brazil. *Brazilian Review of Econometrics*, 12, 223-241.

Luna, I., and Ballini, R. (2011). Top-down strategies based on adaptive fuzzy rule-based systems for daily time series forecasting. *International Journal of Forecasting*, 27, 708-724.

Lütkepohl, H. (1984). Forecasting contemporaneously aggregated vector ARMA Processes. *Journal of Business & Economic Statistics*, 2, 201-214.

Mathews, B., and Diamantopoulos, A. (1986). Managerial intervention in forecasting: an empirical investigation of forecast manipulation. *International Journal of Research in Marketing*, 3, 3-10.

Mathews, B., and Diamantopoulos, A. (1989). Judgmental revision of sales forecasts: a longitudinal extension. *Journal of Forecasting*, 8, 129-140.

Mathews, B., and Diamantopoulos, A. (1990). Judgmental revision of sales forecasts: effectiveness of forecast selection. *Journal of Forecasting*, 9, 407-415.

Mathews, B., and Diamantopoulos, A. (1992). Judgmental revision of sales forecasts - the relative performance of judgementally revised versus non revised forecasts. *Journal of Forecasting*, 11, 569-576.

Meddahi, N., and Renault, E. (2004). Temporal aggregation of volatility models. *Journal of Econometrics*, 119, 355-379.

McCullough, B.D., and Heiser, D.A. (2008). On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics and Data Analysis*, 52, 4570-4578.

Mohammadipour, M., and Boylan, J.E. (2012). Forecast horizon aggregation in integer autoregressive moving average (INARMA) models. *OMEGA: The International Journal of Management Science*, 40, 703-712.

Mohammadipour, M., Boylan, J.E., and Syntetos, A.A. (2012). The application of product-group seasonal indexes to individual products. *Foresight*, 26, 18-24.

- Najafi, M., and Zanjirani Farahani, R. (2014). New forecasting insights on the bullwhip effect in a supply chain. *IMA Journal of Management Mathematics*, 25, 259-286.
- Nikolopoulos, K., and Fildes, R. (2013). Adjusting supply chain forecasts for short-term temperature estimates: a case study in a brewing company. *IMA Journal of Management Mathematics*, 24, 79-88.
- Nikolopoulos, K., and Petropoulos, F. (2015). Forecasting, foresight and strategic planning for black swans. BBSWP/15/0004, Bangor Business School Working Paper Series, Bangor University, UK.
- Nikolopoulos, K., Syntetos, A.A., Boylan, J., Petropoulos, F., and Assimakopoulos, V. (2011). An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis. *Journal of the Operational Research Society*, 62, 544-554.
- Ouwehand, P., van Donselaar, K.H., and de Kok, A.G. (2005). The Impact of the forecasting horizon when forecasting with group seasonal indexes., Working paper 162, Eindhoven University of Technology, The Netherlands.
- Pegels, C. (1969). Exponential forecasting: some new variations. *Management Science*, 15, 311-315.
- Petropoulos, F., Fildes, R., and Goodwin, P. (2015). Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour? *European Journal of Operational Research*, advance online publication, doi: 10.1016/j.ejor.2015.06.002.
- Petropoulos, F., and Kourentzes, N. (2014). Improving forecasting via multiple temporal aggregation. *Foresight*, 34, 12-17.
- Petropoulos, F., and Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66, 914-924.
- Porras, E. M., and Dekker, R. (2008). An inventory control system for spare parts at a refinery: An empirical comparison of different reorder point methods. *European Journal of Operational Research*, 184, 101-132.
- Raghunathan, S. (2001). Information sharing in a supply chain: a note on its value when demand is non-stationary. *Management Science*, 47, 605-610.
- Raghunathan, S. (2003). Impact of demand correlation on the value of and incentives for information sharing in a supply chain. *European Journal of Operational Research*, 146, 634-649.
- Regattieri, A., Gamberi, M., Gamberini, R., and Manzini, R. (2005). Managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management*, 11, 426-431.
- Rostami-Tabar, B., Babai, M.Z., Syntetos, A., and Ducq, Y. (2013). Demand forecasting by temporal aggregation. *Naval Research Logistics*, 60, 479-498.
- Rostami-Tabar, B., Babai, M.Z., Syntetos, A., and Ducq, Y. (2014). A note on the forecast performance of temporal aggregation. *Naval Research Logistics*, 61, 489-500.
- Rostami-Tabar, B., Babai, M.Z., Ducq, Y., and Syntetos, A. (2015). Non-stationary demand forecasting by cross-sectional aggregation. *International Journal of Production Economics*, advance online publication, doi: 10.1016/j.ijpe.2015.10.001.

- Sbrana, G., and Silvestrini, A. (2013). Forecasting aggregate demand: Analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework. *International Journal of Production Economics*, 146, 185-198.
- Shlifer, E., and Wolff, R.W. (1979). Aggregation and proration in forecasting. *Management Science*, 25, 594–603.
- Silvestrini, A., and Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys*, 22, 458-497.
- Singh, S.K. (2013). Chief Operations Officer (COO) of Arkieva (<http://www.arkieva.com>). Private communications to the corresponding author.
- Spithourakis, G.P., Petropoulos, F., Babai, M.Z., Nikolopoulos, K., and Assimakopoulos, V. (2011). Improving the performance of popular supply chain forecasting techniques: an empirical investigation. *Supply Chain Forum: An international Journal*, 12, 16-25.
- Spithourakis, G.P., Petropoulos, F., Nikolopoulos, K., and Assimakopoulos, V. (2014). A systemic view of the ADIDA framework. *IMA Journal of Management Mathematics*, 25, 125-137.
- Spithourakis, G.P., Petropoulos, F., Nikolopoulos, K., and Assimakopoulos, V. (2015). Amplifying the learning effects via a forecasting and foresight support system. *International Journal of Forecasting*, 31, 20-32.
- Strijbosch, L.W.G., Heuts, R.M.J., and Moors, J.J.A. (2008). Hierarchical estimation as a basis for hierarchical forecasting. *IMA Journal of Management Mathematics*, 19, 193-205.
- Strijbosch, L.W.G., Heuts, R.M.J., and van der Schoot, E.H.M. (2000). A combined forecast-inventory control procedure for spare parts. *Journal of the Operational Research Society*, 51, 1184-1192.
- Strijbosch, L.W.G., and Moors, J.J.A. (2010). Calculating the accuracy of hierarchical estimation. *IMA Journal of Management Mathematics*, 21, 303-315.
- Syntetos, A.A. (2014). Forecasting by temporal aggregation. *Foresight*, 34, 6-11.
- Syntetos A.A., and Boylan, J.E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21, 303-314.
- Syntetos, A.A., and Boylan, J.E. (2006). On the stock-control performance of intermittent demand estimators. *International Journal of Production Economics*, 103, 36-47.
- Syntetos, A.A., Boylan, J.E, and Croston, J.D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, 56, 495-503.
- Syntetos, A.A., Boylan, J.E., and Disney, S.M. (2009). Forecasting for inventory planning: a 50-year review. *Journal of the Operational Research Society*, 60 (S1), 149-160.
- Syntetos, A.A., Georgantzias, N.C., Boylan, J.E., and Dangerfield, B.C. (2011). Judgement and supply chain dynamics. *Journal of the Operational Research Society*, 62, 1138 - 1158.

Syntetos, A.A., Kholidasari, I., and Naim, M. (2015a). The effects of integrating management judgement into OUT levels: in or out of context? *European Journal of Operational Research*, advance online publication, doi: 10.1016/j.ejor.2015.07.021.

Syntetos, A.A., Nikolopoulos, K., and Boylan, J.E. (2010). Judging the judges through accuracy-implication metrics: the case of inventory forecasting. *International Journal of Forecasting*, 26, 134-143.

Syntetos, A.A., Nikolopoulos, K., Boylan, J.E., Fildes, R., and Goodwin, P. (2009). The effects of integrating management judgment into intermittent demand forecasts. *International Journal of Production Economics*, 118, 72-81.

Syntetos, A.A., Teunter, R.H., and Prak, D. (2015b). On the calculation of safety stocks. *European Journal of Operational Research*, under review (working paper available from the corresponding author).

Tiao, G.C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, 59, 525-531.

Viswanathan, S., Widiarta, H., and Piplani, R. (2008). Forecasting aggregate time series with intermittent subaggregate components: top-down versus bottom-up forecasting. *IMA Journal of Management Mathematics*, 19, 275-287.

Weatherford, L.R., Kimes, S.E., and Scott, D.A. (2001). Forecasting for hotel revenue management: Testing aggregation against disaggregation. *The Cornell Hotel and Restaurant Administration Quarterly*, 42, 53-64.

Weiss, A.A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, 26, 271-281.

Widiarta, H., Viswanathan, S., and Piplani, R. (2007). On the effectiveness of top-down strategy for forecasting autoregressive demands. *Naval Research Logistics*, 54, 176-188.

Widiarta, H., Viswanathan, S., and Piplani, R. (2008). Forecasting item-level demands: an analytical evaluation of top-down versus bottom-up forecasting in a production-planning framework. *IMA Journal of Management Mathematics*, 19, 207-218.

Widiarta, H., Viswanathan, S., and Piplani, R., (2009). Forecasting aggregate demand: an analytical evaluation of top-down versus bottom-up forecasting in a production planning framework. *International Journal of Production Economics*, 118, 87-94.

Willemain, T. R., Smart, C. N., Shocker, J. H., and DeSautels, P. A. (1994). Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting*, 10, 529-538.

Willemain, T.R., Smart, C.N., and Schwarz, H.F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, 20, 375-387.

Withycombe, R. (1989). Forecasting with combined seasonal indexes. *International Journal of Forecasting*, 5, 547-552.

- Yelland, P. (2010). Bayesian forecasting of parts demand. *International Journal of Forecasting*, 26, 374-396.
- Zellner, A., and Tobias, J. (2000). A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting*, 19, 457-469.
- Zhang, X. (2004). Evolution of ARMA demand in supply chains, *Manufacturing and Service Operations Management*, 6, 195-198.
- Zotteri, G., and Kalchschmidt, M. (2007). A model for selecting the appropriate level of aggregation in forecasting processes. *International Journal of Production Economics*, 108, 74-83.
- Zotteri, G., Kalchschmidt, M., and Caniato, F. (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, 93-94, 479-491.
- Zotteri, G., Kalchschmidt, M., and Sacconi, N. (2014). Forecasting by cross-sectional aggregation. *Foresight*, 35, 35-41.

APPENDIX A: QUESTIONNAIRE

Supply chain forecasting

Distributed to: 26th European Conference on Operational Research (EURO), 1-4 July, 2013

Dear Forecasting researcher & practitioner

We have recently been commissioned by the European Journal of Operational Research (EJOR, Elsevier) to conduct a review article on *Supply Chain Forecasting (SCF)*. We would like to consult as many colleagues as possible and organise the paper around the areas that the forecasting community feels collectively that are important rather than reflect our own background and particular perspectives. This questionnaire aims to identify the most important areas that should be addressed in that paper. The responses are anonymous, unless you would like to indicate your name and participate in follow up discussions during the preparation of the manuscript. In any case the descriptive statistics of your responses will shape the structure of this review article. (Please return the questionnaire either to the session chair who distributed it or to John Boylan.) Thank you very much for your time.

A.A. Syntetos, M.Z. Babai, J.E. Boylan, S. Kolassa, & K. Nikolopoulos

1. Name the 3 most important research areas in SCF during the last 20 years (*with #1 being the most important, everywhere hereafter*)

1.
2.
3.

2. Name the 3 most important academic developments in SCF during the last 20 years

1.
2.
3.

3. Name the 3 most important articles (theoretical or practitioner oriented) in SCF during the last 20 years

1.
2.
3.

4. Name the 3 most important SCF developments for practitioners during the last 20 years

1.
2.
3.

5. Name the 3 most important SCF developments for software developers during the last 20 years

1.
2.
3.

6. What is the 'next big thing' in SCF research on the horizon?

.....

7. What breakthrough could really change the way SCF practitioners do their job in the years to come?

.....

APPENDIX B: SURVEY RESULTS AND DISCUSSION

We have run the questionnaire through a convenience sample of 43 academics; this was intentional as we had to obtain expertise in specific fields of interest, predominantly that of Supply Chain Forecasting (SCF), and as such sampling should be well targeted to probe such expertise. (At this point we should mention that we hoped to include some practitioners amongst our respondents, but unfortunately none were forthcoming.) Sampling has been executed in two phases, initially at the EURO-INFORMS Joint International Meeting in Rome 01-04 July 2013 via face-to-face invitations and since then with an online version of the exact same questionnaire disseminated via the online survey platform *SurveyMonkey*; for that latter phase, there were three main waves of data collection after two sets of direct invitations via emails in the near past, and more recently in October 2014 with an open call via the Oracle – the monthly newsletter of the International Institute of Forecasters (www.forecasters.org).

In the face-to-face invitations, the response rate was close to 100% while in the final phase and the open call it cannot be properly estimated: when invitations were provided via direct emails an average of about 40% of experts have responded to our call to share their views and expertise on the topic. However the great majority of them only partially completed the questionnaire. Nevertheless, as far as the first question is concerned, which is the most important one for the purposes of this review paper, 98% of the participants in this survey provided their views.

Question	Response rate (%)
Q1.1	98%
Q1.2	91%
Q1.3	81%
Q2.1	65%
Q2.2	53%
Q2.3	51%
Q3.1	44%
Q3.2	37%
Q3.3	28%
Q4.1	51%
Q4.2	42%
Q4.5	33%
Q5.1	47%
Q5.2	33%
Q5.3	23%
Q6	49%
Q7	49%

Table B1. Response rate per question (out of the 43 questionnaires)

Given that all the responses are in a narrative format, coding was necessitated so as to present the main findings of this survey and expose the extent to which these reflect our own team's understanding and perceptions of the most important topics and trends in the field of SCF.

Hereafter, we will focus our analysis on the first question "*Name the 3 most important research areas in Supply Chain Forecasting (SCF) during the last 20 years* (with #1 being the most important, everywhere hereafter)" that includes 42, 39 and 35 responses for the first, second and third most important area respectively. In total, 116 qualitative answers were received that were coded as follows:

CODE (alphabetical)	TERMS included in coding
ANALYTICS	Analytics, Big Data
BULLWHIP	Bullwhip
COLLABORATION	Collaboration, Information Sharing
HIERARCHIES	Hierarchical Forecasting, Aggregation
IT/IS	ICT, IS, Systems, Knowledge Management, DBMS
JUDGMENT	Judgment, Integration of Judgment, Changes in Statistical Forecasts, Group Forecasting
METHODS	Methods, Forecasting Methods, Demand Forecasting, Intermittent Demand, GLM, Exponential smoothing, ARIMA, Causal Models, Nowcasting
OTHER	Value Added, Spot Markets, Sales, Marketing, Pricing, Reliability of Information, Process improvement
PLANNING	MRP, Planning, Newspaper Model, Stock control, ECR, Lean, Inventory, Monitoring, Control, Spare Parts, Management, Improvement, Lifecycle Management
SELECTION	Selection of Methods, Automatic Prediction, Evaluation, Accuracy
SUPPLY CHAIN	Length of Supply Chain, Global Supply chains
SUSTAINABILITY	Sustainability, Green Supply Chains
UNCERTAINTY	Error, Uncertainty, Special Events, Interruptions, Demand Versus Sales

Table B2. Response CODEings

In the following Table (B3) we summarise the responses per coding and per preference (1st, 2nd or 3rd) from the experts in our sample.

CODE	Frequency of response		
	As Most Important (1 st)	Second (2 nd)	Third (3 rd)
ANALYTICS	1	1	1
BULLWHIP	2	3	1
COLLABORATION	8	8	4
HIERARCHIES	-	1	1
IT/IS	3	1	1
JUDGMENT	4	3	4
METHODS	6	7	7
OTHER	1	1	5
PLANNING	6	8	3
SELECTION	3	1	2
SUPPLY CHAIN	2	2	-
SUSTAINABILITY	1	-	2
UNCERTAINTY	5	3	4
Total	(42)	(39)	(35)

Table B3. Responses rate per CODE (Grey-shaded: the two most popular - including ties)

Furthermore our CODEs have been classified both in terms of their relevance to *single-echelon* or *across-supply-chain* topics, as well as according to the extent to which they are ‘unambiguously’ *relevant* to supply chain forecasting or not; in both cases (collective – across the members of our team) academic judgment needed to be exercised. This resulted in Table B4.

CODE (alphabetical)	CATEGORISATION	
CODE (alphabetical)	Single-Echelon Topic or <u>Across-Supply-Chain</u>	<u>Relevance</u> of Topic to SCF
ANALYTICS	Supply Chain	High
BULLWHIP	Supply Chain	High
COLLABORATION	Supply Chain	High
HIERARCHIES	Both	High
IT/IS	Both	Medium
JUDGMENT	Both	High
METHODS	Single-Echelon	High
OTHER	Both	Low
PLANNING	Both	Medium
SELECTION	Single-Echelon	High
SUPPLY CHAIN	Supply Chain	Low
SUSTAINABILITY	Supply Chain	Low
UNCERTAINTY	Both	High

Table B4. Categorisation of CODEings

We then decided to focus on the codes (and related keywords) that relate to either across-supply-chain topics or both single-echelon and across-supply chain topics as well as being highly relevant to FORECASTING, resulting in Table B5.

CODE	Frequency of response			
	As Most Important	Second	Third	Total Ranking
ANALYTICS	1	1	1	6
BULLWHIP	2	3	1	13
COLLABORATION	8	8	4	44
HIERARCHIES	-	1	1	3
JUDGMENT	4	3	4	22
UNCERTAINTY	5	3	4	25
Revised Total	(20)	(19)	(15)	<i>(3 points for 'Most Important', 2 points for 'Second' and 1 point for 'Third')</i>

Table B5. Responses focusing on SCF per CODE

Furthermore we wanted to see how our codes and the respective keywords relate to our methodological three-dimensional paradigm of *length*, *depth* and *time*. Given our definitions in Section 1 we believe that 'BULLWHIP' and 'COLLABORATION' clearly fall within the *length* dimension, 'HIERARCHIES' within *depth*, while 'JUDGMENT' has strong connections to *time*. More discussion could be brought into where 'ANALYTICS' do actually fit in, but our consensus was that descriptive, predictive and prescriptive analytics primarily attempt to identify optimal horizons, frequencies, points in time, historical data to be used, etc., and as such this mostly relates to the dimension of *time* rather than anything else. This classification is also consistent with the classification of 'JUDGMENT', as analytics and judgment are supposed to complement each other in the decision making spectra. Finally the most

difficult one to categorise was 'UNCERTAINTY' as this includes both keywords like '*interruptions*' and '*demand vs. sales*' that link more to the dimension of *length*, as well as keywords like '*errors*' and '*special events*' that relate more to the dimension of *time*. Therefore, we decided to equally break this category into the respective two methodological dimensions of *length* and *time*. All this analysis resulted in our final Table B6 and the respective Figure B1.

<i>Dimension</i>	CODE	CODE score	<i>Dimension score</i>
<i>length</i>	BULLWHIP	13	69.5
	COLLABORATION	44	
	UNCERTAINTY	12.5	
12.5		40.5	
<i>time</i>	ANALYTICS		6
	JUDGMENT		22
<i>depth</i>	HIERARCHIES	3	3

Table B6. Linking the Experts' survey CODEs with the methodological *dimensions of length, depth and time*

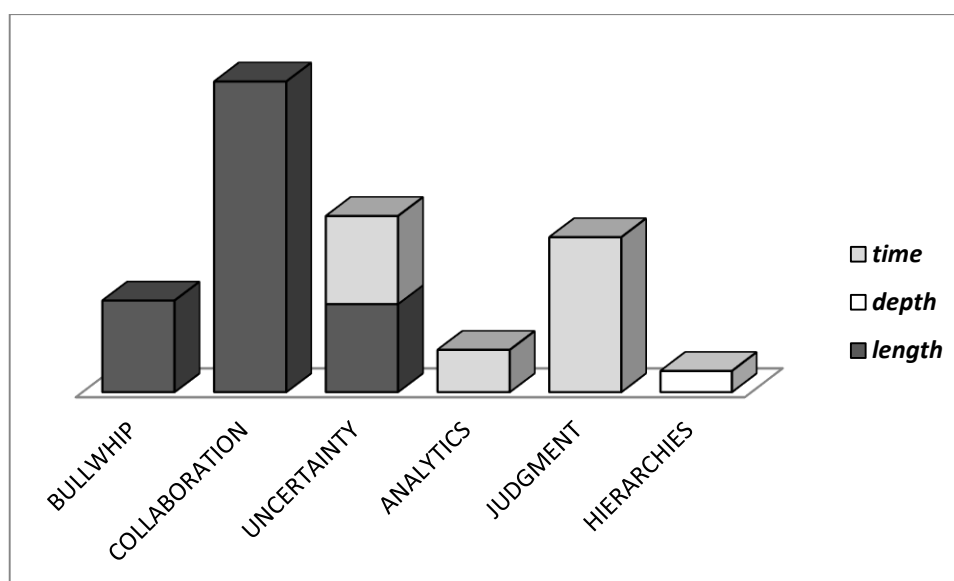


Figure B1. Graphical representation of Experts' perceived importance of SCF topics

This graph has informed our review, combined with what we consider to be the conceptual underpinnings of the major developments in recent years and aligned with our three-dimensional methodological framework of *length*, *depth* and *time*. It is clearly illustrated that the experts see the themes of **Collaboration**, **Uncertainty** and **Judgment** and respectively the dimensions of ***length*** and ***time*** being the most important ones, and as such those should receive more attention in the years to come from researchers in the field.

APPENDIX C: UPSTREAM PROPAGATION OF ARIMA DEMAND PROCESSES AND DOWNSTREAM DEMAND INFERENCE

We consider a downstream supply chain stage that faces an ARIMA (p, d, q^{dns}) demand process and uses a periodic order-up-to-level policy to control its inventory level. The lead-time is denoted by L .

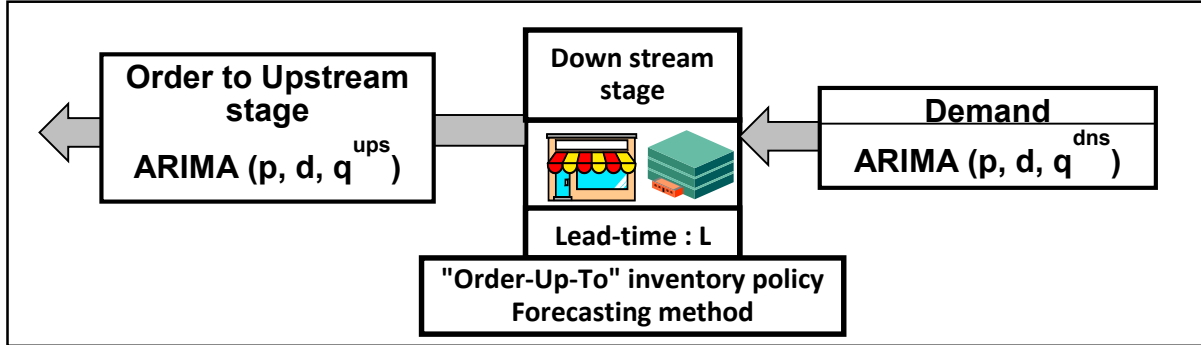


Figure C1. Propagation of demand processes in supply chains under the ARIMA framework

Table C1 shows the relationship between the demand and the order process parameters for a general ARIMA (p, d, q^{dns}) process. The particular results when the downstream stage faces an AR(1) demand process are summarised in Table C2. The results are provided for three possible forecasting methods: i) a minimum mean square error (MMSE), ii) a Single Moving Average (SMA) of length n and, iii) a Single Exponential Smoothing (SES) method. The relevant references to the papers that established the results are also presented.

Downstream demand process	Forecasting method	Upstream demand process	Relationship	Reference
ARIMA(p, d, q^{dns})	MMSE	ARIMA (p, d, q^{ups})	$q^{ups} = \max \{p+d, q^{dns}-L\}$	Gilbert (2005)
	SMA(n)	ARIMA (p, d, q^{ups})	$q^{ups} = q^{dns} + n$	Ali and Boylan (2012)
	SES	ARIMA (p, d, q^{ups})	$q^{ups} \approx t - 1$ where t is the current time period	

Table C1. General results on the propagation of ARIMA demand processes

Ali and Boylan (2012) have shown that the approximate equality of the upstream demand process moving average parameter when SES is used is due to the presence of an extra term in the process expression. However, this term becomes negligible rapidly for increasing finite values of t .

Downstream demand process	Forecasting method	Upstream demand process	Reference
AR (1)	MMSE	ARMA(1,1)	Alwan et al. (2003) ¹
	SMA(n)	ARMA(1, n)	
	SES	ARMA(1,?)	

Table C2. Results on the propagation of an AR(1) demand process

Based on the results provided by Gilbert (2005), Ali and Boylan (2011) have analysed the approach referred to as *Downstream Demand Inference* (DDI) whereby the upstream stage in a supply chain can mathematically infer the demand information from orders received from the downstream stage. They have considered a supply chain where an upstream stage with lead time L faces an $ARIMA(p, d, q^{ups})$ demand process. They have shown that the inference of the downstream demand process is feasible depending on the forecasting method used by the downstream stage and the properties of the demand process. The results are summarised in Table C3.

Upstream demand process	Forecasting method	DDI	
		Feasibility	Condition
$ARIMA(p, d, q^{ups})$	MMSE	YES	$p + d < q^{ups}$
	SMA(n)	YES	n is known
	SES	NO	

Table C3. DDI feasibility for an $ARIMA(p, d, q^{ups})$ demand process

To illustrate the results given in Table C3, if the demand process faced by the upstream stage is an $ARIMA(1,1,2)$ process and the lead time is equal to 1, DDI is not feasible if a MMSE forecasting method is used as this demand process could propagate from various processes at the downstream stage, namely: $ARIMA(1,1,2)$, $ARIMA(1,1,1)$ and $ARIMA(1,1,0)$. However, DDI is feasible if a SMA(3) forecasting method is used and the inferred downstream demand is an $ARIMA(1,1,0)$. Furthermore, if for example the demand process faced by the upstream stage is an $ARIMA(1,1,3)$, the demand at the downstream stage can be inferred as an $ARIMA(1,1,4)$ under a MMSE forecasting method and an $ARIMA(1,1,0)$ under a SMA(3) forecasting method.

¹Alwan et al. (2003) assume that an infinite demand history is considered to derive these results.

APPENDIX D: HIERARCHICAL DEMAND AGGREGATION

Hierarchical (or cross-sectional or contemporaneous) aggregation occurs when the aggregation takes place across a number of SKUs at one specific time period (Silvestrini and Veredas, 2008). It has been shown in the academic literature that the cross-sectional aggregation of independent ARIMA processes result also in an ARIMA process (Box and Jenkins, 1970; Anderson, 1975; Granger and Morris, 1976). In particular:

- The cross-sectional aggregation of N independent ARMA processes is an ARMA process, where:

$$\sum_{j=1}^N ARMA(p_j, q_j) = ARMA(x, y), \text{ where } x \leq \sum_{j=1}^N p_j \text{ and } y \leq \max(x - p_j + q_j) \text{ for any } j = 1..N$$

- The cross-sectional aggregation of two independent ARIMA processes is an ARIMA process, where:

$$\sum_{j=1}^2 ARIMA(p_j, d_j, q_j) = ARIMA(p, d, q), \text{ where}$$

$$p \leq p_1 + p_2 \text{ and } d = \max(d_1, d_2), \text{ and}$$

$$\begin{cases} q \leq \max(p_1 + q_2 + d_1 - d_2, p_2 + q_1) & \text{if } d_1 \geq d_2 \\ q \leq \max(p_2 + q_1 + d_2 - d_1, p_1 + q_2) & \text{if } d_2 \geq d_1 \end{cases}$$

Examples of ARIMA type cross-sectional aggregated processes are illustrated in Table D1 (for 2 non-aggregated demand processes). Note that ϕ_i denotes the auto-correlation parameter of the i^{th} process (ϕ_1 for the 1st process and ϕ_2 for the 2nd process).

Non-aggregated demand process	Cross-sectional aggregated process	Reference
MA(1) + MA(1)	MA(1)	Box and Jenkins (1970)
AR(1) + AR(1) with $\phi_1 \neq \phi_2$ and $\phi_1 \neq -\phi_2$	ARMA(2,1)	Granger and Morris (1976)
ARMA(1,1) + ARMA(1,1) with $\phi_1 = \phi_2$	ARMA(1,1)	
AR(1) + AR(1) with $\phi_1 = \phi_2$	AR(1)	Harvey (1993).
AR(1) + AR(1) with $\phi_1 = -\phi_2$	AR(2)	
ARIMA(0,1,1) + ARMA(0,1,1)	ARIMA(0,1,1)	Silvestrini and Veredas (2008)

Table D1. Example of the relationship between non-aggregated and cross-sectionally aggregated ARIMA processes

Existing approaches to cross-sectional forecasting usually involve either a bottom-up (BU) or a top-down (TD) approach (or a combination of both). When forecasting at the aggregate level is performed, BU involves the aggregation of individual SKU forecasts to the group level whereas TD relates to forecasting directly at the group level (i.e. first aggregate requirements and then extrapolate directly at the aggregate level). When forecasting at the subaggregate level is of interest, BU relates to direct extrapolation at the subaggregate level whereas TD involves the disaggregation of the forecasts produced at the group level.

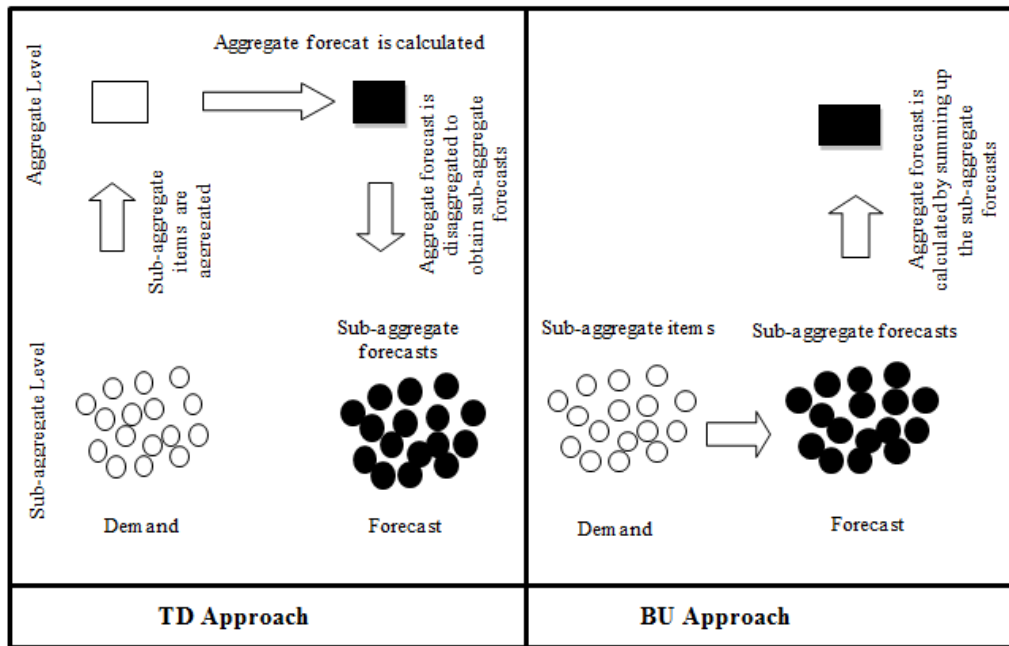


Figure D1. Schematic diagram of TD and BU forecasting approaches

As illustrated in Figure D1, these approaches work as follows: The TD approach consists of the following steps: i) subaggregate demand items are aggregated; ii) the forecast of aggregate demand is produced by applying a forecasting method at the aggregate level, and iii) the forecast is disaggregated back to the original level by applying an appropriate disaggregation method (often based on historical proportions), if a subaggregate forecast is needed. In the BU approach: i) subaggregate demand forecasts are produced directly for the subaggregate items; ii) the aggregate forecast is obtained by combining individual forecasts for each SKU, i.e. potentially a separate forecasting model may be used for each item in the product family (Zotteri et al., 2005).

An important issue that has attracted the attention of many researchers as well as practitioners over the last few decades is the comparative forecast performance of such cross-sectional forecasting approaches. The relevant work has been conducted by means of real data as well as ARIMA-type demand processes. Table D2 presents a summary of the literature that has analysed the comparative performance of the two cross-sectional approaches discussed above along with a brief description of the assumptions made in terms of the analysed demand and forecasting methods.

Recently, Hyndman et al. (2011) have proposed a new hierarchical forecasting approach that provides optimal forecasts across all the levels of a hierarchy. Based on a simulation study using Australian tourism demand, they have shown that the proposed approach provides more accurate forecasts than the TD and BU approaches. One useful feature of the proposed method is that the forecasts may come from any forecasting model.

Reference	Analysed demand	Forecasting method
Barnea and Lakonishok (1980)	Corporate accounting data from 5 companies	Single exponential smoothing
Lutkepohl (1984)	k-dimensional ARMA(p,q) demand process	Maximum likelihood procedure
Gross and Sohl (1990)	Data of 18 series from US Gypsum, Inc.	Exponential smoothing (single, double, Holt, Winters, and quadratic), regression (linear, quadratic and logarithmic)
Dangerfield and Morris (1992)	Data of 178 series from the M-competition	Winters triple exponential smoothing
Gordon et al. (1997)	Sales data on 15,000 products	Triple exponential smoothing
Fliedner (1999)	MA(1) process	Single exponential smoothing and Simple Moving Average
Zellner and Tobias (2000)	Data on the GDP growth rates from 18 countries	Auto-regression of order three with lagged leading indicator input variables
Weatherford et al. (2001)	Data from Marriott for two large business hotels	Classical-pickup method, moving averages, linear regression, simple exponential smoothing and random walk.
Brännäs et al. (2002)	INAR(1) and INMA(1) processes	Conditional expectation
Widiarta et al. (2007)	AR(1) process	Single exponential smoothing
Widiarta et al. (2008, 2009)	MA(1) process	Single exponential smoothing
Viswanathan et al. (2008)	Theoretically generated intermittent demand data	Croston's method
Sbrana and Silvestrini (2013)	ARIMA(0,1,1) process	Multivariate exponential smoothing framework
Rostami-Tabar et al. (2015)	ARIMA(0,1,1) process and demand of 424 time series from the retailing sector	Single exponential smoothing

Table D2. Literature that analyses the TD and BU approaches

APPENDIX E: TEMPORAL AGGREGATION OF DEMAND

Temporal aggregation (or aggregation across time) refers to the process by which a low frequency time series (e.g. monthly) is derived from a high frequency time series (e.g. weekly) (Nikolopoulos et al., 2011). There are two different types of temporal aggregation: non-overlapping and overlapping. In the former case, the time series are divided into consecutive non-overlapping buckets of time where the length of the time bucket equals the aggregation level. The latter case is similar to a moving window technique where the window's size equals the aggregation level. At each period, the window is moved one step ahead, and so the oldest observation is dropped and the newest is included. Figure E1 illustrates the case of temporally aggregating demand from weekly to monthly data, i.e. the aggregation level is equal to 4.

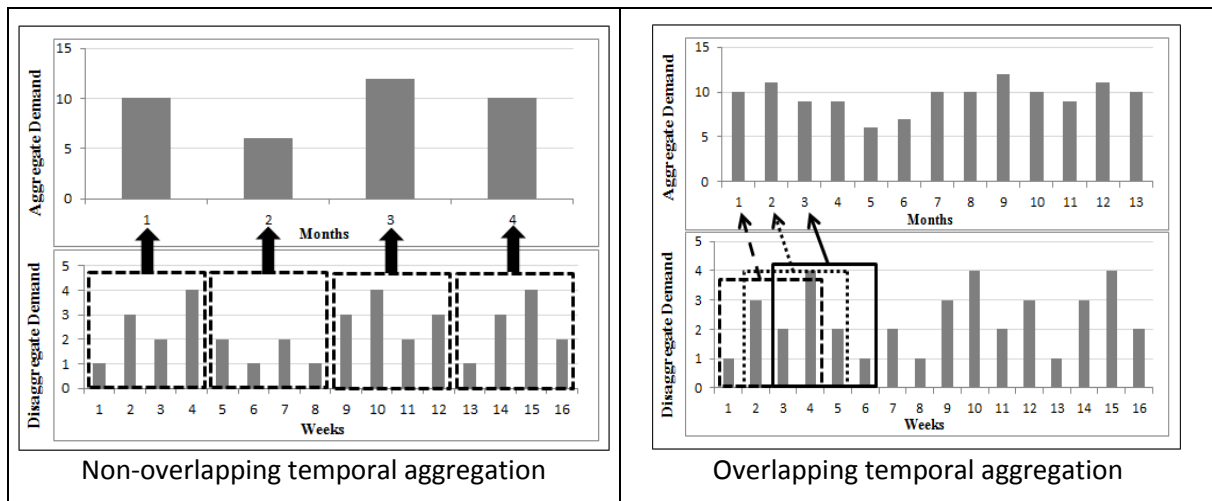


Figure E1. Temporal aggregation (from weekly to monthly data)

Note that the information loss in the overlapping case is negligible as compared to the non-overlapping one. This is an important observation in terms of data availability and for the cases where little history of data is available.

Weiss (1984) has shown that the non-overlapping temporal aggregation of an ARIMA (p, d, q) process is an ARIMA (p, d, r) demand process where $r = \left[\frac{(p(m-1) + (d+1)(m-1) + q)}{m} \right]$ with m denoting the aggregation level and $[X]$ the integer part of X . Examples of the non-overlapping aggregation of ARIMA type processes are provided in Table E1.

Non-aggregated demand process	Non-overlapping temporal aggregated process
AR(1)	ARMA(1,1)
MA(1)	MA(1)
ARMA(1,1)	ARMA(1,1)
ARIMA(0,1,1)	ARIMA(0,1,1)

Table E1. Examples of the relationship between non-aggregated and non-overlapping temporally aggregated ARIMA processes

Luiz et al. (1992) have shown that the overlapping temporal aggregation of an ARIMA (p, d, q) process is an ARIMA (P, d, Q) demand process where $P \leq p$ and $Q \leq q + m - 1$ (and where m denotes the aggregation level).

Some work has also been conducted to characterise the overlapping temporal aggregation of integer autoregressive moving average (INARMA) demand processes. Mohammadipour and Boylan (2012) have shown that the overlapping temporal aggregation over m periods of an INARMA(p, q) process with Poisson innovation terms (with mean λ) results in an INARMA(p, q) process with the same autoregressive and moving average parameters and the innovation parameter of $m \lambda$. In Table E2, we show some results that have been developed under the assumption of Poisson distributed demand innovation terms.

Non-aggregated demand process	Overlapping temporal aggregated process	Reference
INAR(1)	INARMA(1,1)	Brännäs (2002)
INMA(1)	INMA(1)	
INARMA(p, q)	INARMA(p, q)	Mohammadipour and Boylan (2012)

Table E2. Example of relationship between non-aggregated and non-overlapping temporal aggregated INARMA processes

During recent years, there has been an increasing interest in analysing the forecasting performance of both overlapping and non-overlapping temporal aggregation approaches and very interesting analytical and empirical results have been presented especially under the ARIMA and INARMA demand framework. Table E3 presents a summary of the relevant references and a brief description of their assumptions in terms of the aggregation approach, the analysed demand data and the forecasting methods being used.

Reference	Aggregation approach	Analysed demand	Forecasting method
Amemiya and Wu (1972)	Non-overlapping	AR(p) process	Three linear predictors
Tiao (1972)	Non-overlapping	IMA(d, q) process	Conditional expectation
Wei (1979)	Non-overlapping	Seasonal-type ARIMA processes and data from US civilian workers	Conditional expectation
Willemain et al. (1994)	Non-overlapping	Demand from 16 empirical data sets	Croston's method
Brännäs et al. (2002)	Overlapping	INAR(1) and INMA(1) processes	Conditional expectation
Porras and Dekker (2008)	Overlapping	Demand of 14,000 spare parts from a petrochemical company	Bootstrapping method
Athanasopoulos et al. (2011)	Non-overlapping	Data of 366 time series from the tourism industry	Innovations state space models for exponential smoothing, ARIMA methodology, damped trend, theta method and naïve
Luna and Ballini (2011)	Non-overlapping	Data of 111 time series related to cash money withdrawals	Neural networks methods

Nikolopoulos et al. (2011)	Non-overlapping	Demand of 5,000 SKUs from the Royal Air Force (RAF, UK)	Naïve, Croston and Syntetos-Boylan Approximation (SBA)
Spithourakis et al. (2011)	Non-overlapping	Data of 1,428 series from the M3-Competition	Naïve, SES, Theta, Holt and damped forecasting methods
Mohammadipour and Boylan (2012)	Overlapping	INARMA(p,q) and 16,000 SKUs from the Royal Air Force (RAF, UK)	Conditional expectation
Babai et al. (2012)	Non-overlapping	Demand of 5,000 SKUs from the Royal Air Force (RAF, UK)	SES, Croston and SBA
Rostami-Tabar et al. (2013, 2014)	Non-overlapping	AR(1), MA(1), ARMA(1,1) processes and 1,798 time series from the retailing sector	Single exponential smoothing
Kourentzes et al. (2014)	Non-overlapping	Demand of 3,000 time series from the M3-Competition	Exponential smoothing methods (ETS method)

Table E3. Literature dealing with temporal aggregation approaches